STRUCTURAL AND BIOCHEMICAL STUDIES OF ENZYMES INVOLVED IN THE BIOSYNTHESIS OF VALUE-ADDED PRODUCTS

BY

NEKTARIA PETRONIKOLOU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Satish K. Nair, Chair
Professor John E. Cronan
Professor Huimin Zhao
Assistant Professor Hong Jin

ABSTRACT

Enzymes have been used for decades for the industrial production of high value chemicals such as food additives, antibiotics and other pharmaceutical products. However, in recent years the significant advances in biocatalyst development have led to a rise in the number of commodity chemicals being produced by enzymes instead of extraction from plants or chemical synthesis. Such products include biodiesel and wax esters.

For this industrial exploitation of enzymes, it is vital to have an extensive understanding of how the desired enzymes work, including what the enzymes' substrate preferences and stability are. This knowledge comes from the structural and biochemical characterization of these biocatalysts, and this is the scope of this dissertation; to broaden our knowledge and understanding of enzymes that are involved in the biosynthesis of value-added products, such as biodiesel and drugs.

Biodiesel is a mixture of long chain fatty acid alkyl esters (FAAEs), produced mainly by the transesterification of fatty acids derived from vegetable oils and animal fats. In recent years, biodiesel has emerged as a viable resource for utilization in green energy production. However, current production processes utilize methods that are cost ineffective and generate waste byproducts. Recent progress in the field makes use of biocatalysts that are effective in diverting products from primary metabolism to yield fatty acid methyl and ethyl esters in bacterial systems (as well as other commodity chemicals). Two of these enzymes have been the focus of my study: a fatty acid O-methyltransferase (FAMT) from *Mycobacterium marinum* (MmFAMT), and a bifunctional wax ester synthase/diacylglycerol acyl-transferase (WS/DGAT) from *Marinobacter aquaeolei* VT8 (Ma-WS/DGAT).

In Chapter 1, I present the first kinetic and structural characterization of a mycobacterial fatty acid *O*-methyltransferase (MmFAMT) previously utilized for the production of biodiesel in *E. coli*. The structure revealed unexpected similarity with methyltransferases from plant natural product metabolism, and an active-site cavity that can accommodate fatty acids up to 12-14 carbons long. Our results provide the framework for further optimization of an *in vivo* system that utilizes MmFAMT for production of biodiesel.

In Chapter 2, I discuss the structure of a bifunctional wax ester synthase/diacylglycerol acyl-transferase (Ma-WS/DGAT), which is to our knowledge the first structure for any member of this enzyme family to date. Wax esters (WE) are esters of long chain fatty acids and long chain alcohols and have widely been used in cosmetics, lubricants, candles and the food industry. In addition, WE that consist of medium chain fatty acids and ethanol (fatty acid ethyl esters, FAEEs) are biodiesel. Today, there is a strong demand for the development of a large-scale process for the production of low cost WE. The structure may facilitate future engineering efforts aiming at the optimization of this biocatalyst towards production of desired wax esters for industrial use.

In addition to biodiesel, value-added chemicals produced by enzymes include bioactive compounds such as chemotherapeutics and antibiotics. In Chapters 3 and 4, I present structural and biochemical data that expand our knowledge for enzymes involved in the biosynthesis of such compounds.

In Chapter 3, I discuss the molecular basis for the substrate specificity of loganic acid *O*-methyltransferase from *Catharanthus roseus* (CraLAMT), an enzyme that methylates with high stereospecificity an iridoid glycoside to produce loganin. Loganin is an intermediate in the biosynthesis of indole alkaloids in plants - a group of plant natural products with several medicinal uses, and has attracted interest as a bioactive compound itself. Our results may facilitate future

engineering efforts to diversify the CraLAMT substrate scaffold towards the design of new potential bioactive compounds.

In Chapter 4, I present structural and biochemical data for enzymes that are involved in the biosynthesis of phosphonates. Phosphonates are natural products containing a carbon-phosphorous (C-P) bond. These C-P containing molecules inhibit enzymes which utilize phosphate esters and carboxylic acids by mimicking their substrates. At a time when drug-resistant pathogens pose serious threat to human health, phosphonate natural products provide a pool of potential therapeutic candidates. Two such phosphonate containing compounds are rhizocticin and plumbemycin.

Rhizocticin and plumbemycin are di- or tripeptides that contain the same phosphonic amino acid at their C-termini, but differ at their N-termini. This unusual amino acid is (Z)-L-2-amino-5-phosphono-3-pentenoic acid (APPA), and is the active part of rhizocticin and plumbemycin. Upon entering the cell, host peptidases release APPA which subsequently inhibits threonine synthase (ThrC) and kills the cell. Because *thrC* is an essential gene in bacteria and fungi, but is absent in humans, rhizocticin and plumbemycin are attractive as potential antimicrobials. In Chapter 4, I present data that provide insight into the mechanism of inhibition of ThrC by APPA. In addition, our data shed light on how the producing organisms overcome APPA toxicity and attain resistance.

To my beloved husband, Enrique. Without his love and support I wouldn't have gotten through this doctorate.

ACKNOWLEDGEMENTS

I would like to thank all those who have helped me throughout my doctoral studies. The work presented in this dissertation would not have been possible without their support and encouragement.

First and foremost I want to express my sincere gratitude to my advisor Prof. Satish Nair for taking me into his lab and giving me the opportunity to work on interesting projects. I thank him for his patience and guidance, and for giving me great freedom and independence, allowing me to grow as a scientist.

Besides my advisor, I would like to thank the members of my doctoral committee: Prof. John Cronan, Prof. Huimin Zhao and Prof. Hong Jin for taking time from their busy schedule to attend my presentations and for providing me with valuable feedback.

My sincere thanks to my collaborators Prof. Mary Schuler, Prof. William Metcalf and Dr. Manuel Ortega, and all the members of the Mining Microbial Genomes (MMG) Research Theme at the Carl R. Woese Institute for Genomic Biology. Our interactions and stimulating discussions have morphed me into the scientist that I am today.

I would also like to thank all those who have trained and assisted me throughout the years. Special thanks to Dr. Yue Hao for patiently training me and teaching me the fundamentals of protein crystallography. My sincere appreciation to the beamline scientists of LS-CAT at Argonne National Laboratory for assistance with data collection and processing. I want to especially thank Prof. Zdzislaw Wawrzak and Dr. Giovanni Gonzalez-Gutierrez for teaching me how to process my data. I greatly appreciate their kindness and willingness to help me.

I thank previous and current members of the Nair group for stimulating discussions and for making more tolerable, and sometimes even fun, the sleepless nights we worked together collecting data. I especially thank them for creating a positive and fun work environment. Many of these people will always have a special place in my heart. They have been not only my co-workers, but my friends, my “partners in crime”, my support system.

I also want to thank my family and friends whose continuous encouragement and support have helped me navigate through this exciting, but also long and emotionally draining journey. Most importantly, I want to thank my husband, Enrique. His love and emotional support gave me the strength to get up whenever I fell and to keep going. Now that I have reached my destination, I am excited about the new journeys that I am about to embark on, and I couldn’t have asked for a better travel companion. Thank you.

TABLE OF CONTENTS

CHAPTER 1: STRUCTURAL AND BIOCHEMICAL STUDIES OF A MYCOBACTERIAL FATTY ACID METHYLTRANSFERASE[1]

1.1 Introduction

The increasing demand for energy, coupled with the need for energy independence and security, have been the main forces driving the growing interest in the use of biofuels as a potential sustainable energy resource. The Energy Independence and Security Act of 2007 (EISA) mandates increased production of biofuels with the aim of reducing the national dependence on unstable foreign suppliers, as well as reducing pollutants produced by the consumption of fossil fuels [1]. Although biofuels hold great promise, the current methodologies used for their production require large quantities of water, fertilizers and pesticides rendering biofuels less environment-friendly and sustainable than desired [2].

In recent years, biodiesel has emerged as a viable resource for utilization in green energy production. In addition to favorable properties such as biodegradability and low toxicity, biodiesel has properties similar to those of petro-diesel, and can be used as a motor fuel without major engine modifications [3]. It is a mixture of long chain fatty acid alkyl esters (FAAEs), produced mainly by the transesterification of fatty acids derived from vegetable oils and animal fats [4]. These

[1] This chapter is adapted from the following published article:

Nektaria Petronikolou & Satish K. Nair (2015). Biochemical Studies of Mycobacterial Fatty Acid Methyltransferase: A Catalyst for the Enzymatic Production of Biodiesel. *Chemistry & Biology*, 22 (11) 1480-1490.

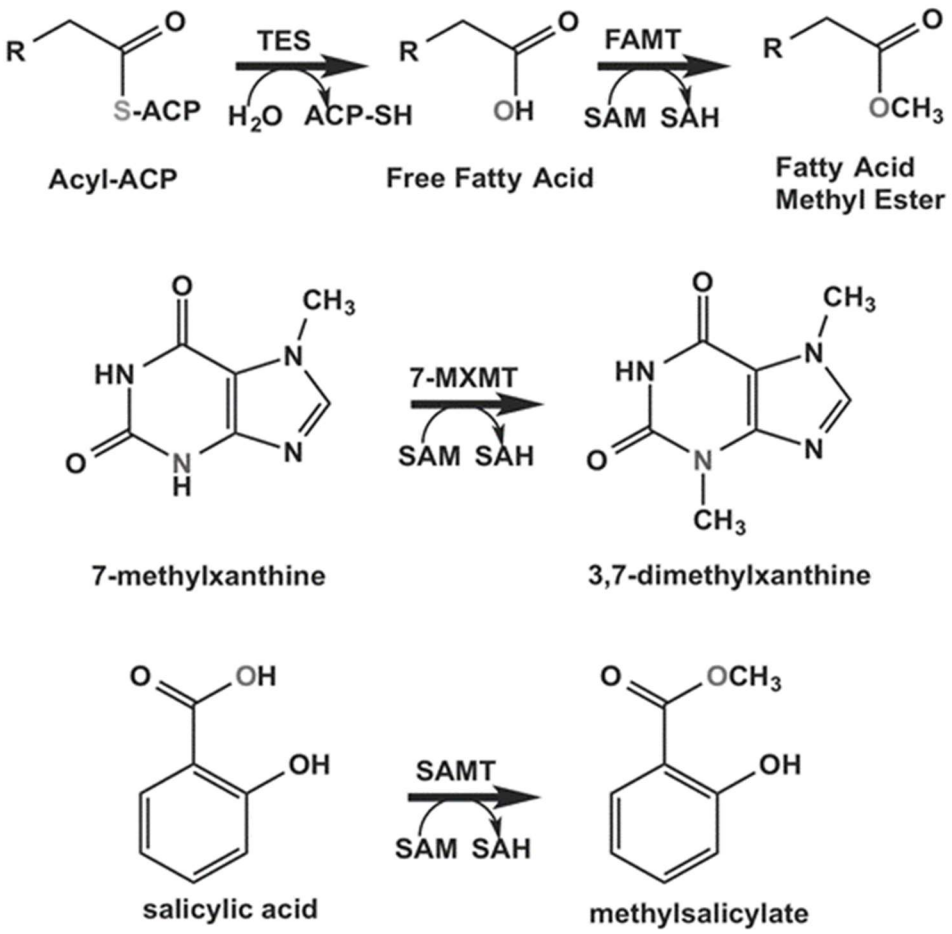

Figure 1. 1: Reactions catalyzed by the fatty acid O-methyltransferase (FAMT) and two of its structural homologs from the SABATH family (7-MXMT: 7-methylxanthine N-methyltransferase, SAMT: salicylic acid O-methyltransferase). Free fatty acids can be released by the action of a thioesterase (TES) on acyl-ACP (ACP: Acyl Carrier Protein).

biodiesel precursors are esters of three fatty acids with glycerol, known as triglycerides (TGs), which can undergo transesterification with an alcohol in the presence of a catalyst to yield FAAEs and glycerol as final products. However, this process is expensive and generates waste, negating the financial and environmental benefits of using biodiesel [5].

With the goal of formulating less energy intensive and more environment-friendly methods for biodiesel production, several laboratories have engineered microorganisms for overproduction of fatty acids [6-9]. More recently, *E. coli* has been engineered to directly produce fatty acid methyl

esters (FAMEs) and 3-hydroxy fatty acid methyl esters (3-OH FAMEs) by using a fatty acid *O*-methyltransferase (FAMT) from *Mycobacterium marinum* (MmFAMT) (Figure 1. 1) [9]. Although the yield of FAMEs produced by this pathway was lower than the maximal FAAEs yields reported [7, 9], the use of FAMT holds promise for *in vivo* biodiesel production.

The fatty acid *O*-methyltransferases are S-adenosylmethionine (SAM) dependent enzymes that catalyze the transfer of a methyl group to the carboxyl group of a fatty acid acceptor [10]. As SAM dependent methyltransferases have important implications for health, the reaction mechanism and structure of these enzymes are well studied and characterized [11]. Mycobacteria specifically possess a large number of SAM dependent methyltransferases that catalyze the modification of mycolic acids, the very long fatty acids found in mycobacterial cell walls [12]. These enzymes can be broadly classified with regards to the identity of the methyl acceptor, and include the cyclopropane fatty acid synthases (which catalyze *C*-methylation of a *cis* double bond) [13], and the *O*-methyltransferases that catalyze methylation on the hydroxyl moiety of phthiocerol, phenolphthiocerol, or ketomycolate [14, 15]. There are three cyclopropane fatty acid synthases with different regiospecificities; CmaA1 catalyzes the installation of the distal cyclopropane ring on α-mycolic acid when overexpressed [16, 17], PcaA installs that proximal cyclopropane ring in α-meroacid [18], and CmaA2 catalyzes the *cis*-cyclopropane synthesis in methoxy-mycolates [15]. Additional examples include MmaA4 that introduces a methyl branch, together with an adjacent hydroxyl group during the formation of both keto- and methoxymycolates [19]. In contrast to these other SAM-dependent enzymes, little is known about fatty acid *O*-methyltransferases (FAMTs) [9, 20-23]. While they were first discovered in Mycobacteria in 1970 [20], there are very few studies published about this class of

methyltransferases [9, 20, 22], and their physiological role in Mycobacteria has yet to be elucidated.

The identification of the mycobacterial FAMTs was facilitated by the sequence analysis of mycobacterial proteins, which suggested that the structure contains Rossman-like fold found in class I methyltransferases [24]. The primary sequences of the mycobacterial enzymes also suggest weak, but notable, similarities to plant natural product methyltransferases, such as dimethylxanthine methyltransferase, (PDB code 2EFJ) [25], salicylate methyltransferase (PDB code 1M6E) [26], as well as other members of a recently characterized protein family classified as the SABATH methyltransferase (Pfam03492 [27]; so named for the first characterized members of this fold class) (Figure 1. 1). However, these plant enzymes are not known to catalyze methylation of fatty acid substrates, and the sequence identity between the mycobacterial enzymes and the plant *O*-methyltransferases is less than 25%.

In order to understand the basis for this biotechnologically relevant enzyme, we carried out biochemical and biophysical characterization of the MmFAMT that has been previously utilized for the *in vivo* production of biodiesel [9]. Cocrystal structures of the enzyme with a variety of bound substrates allow for the identification of features that bestow specificity to fatty acid substrates. Kinetic characterization of active site variants, as deduced from the structural data, facilitates the assignment of roles for various residues and the proposal for a reaction mechanism consistent with the data. These studies should provide a framework for future engineering experiments aimed at adapting FAMTs for *in vivo* high-yield production of biodiesel from fatty acid precursors.

1.2 Materials and Methods

Chemicals and Reagents

All chemical reagents were purchased from Sigma-Aldrich unless otherwise noted. All of the material used for protein production and purification were purchased from GE Healthcare.

Cloning and Site-Specific Mutagenesis

MmFAMT (GenBank accession number NC010612 – gene MMAR3356) was amplified using the PCR with template genomic DNA of *Mycobacterium marinum* strain M (ATCC BAA-535) and primers designed based on the published sequence. The gene was inserted into the NdeI/BamHI sites of a pET28b vector, and this plasmid was used as template for the generation of the site-specific mutants by PCR (Table A. 1). The integrity of all recombinant plasmids was confirmed by sequencing (ACGT, Inc.).

Protein Expression and Purification

Expression vectors bearing wild type or mutant MmFAMT were transformed into *E. coli* Rosetta 2(DE3) cells for heterologous protein production. A 5 ml starter culture was inoculated in 2L Luria-Bertani growth medium supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol. The culture was grown at 37 °C until the absorbance at 600 nm reached 0.6 – 0.8, at which point protein production was induced by addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was then cooled to 18 °C and grown for an additional 18 h. Cells were collected by centrifugation, resuspended in 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 10% glycerol buffer, and lysed by multiple passages through a C5 Emulsiflex (Avestin) cell homogenizer. Following centrifugation of the lysate, the supernatant was applied to a 5 ml

His-Trap (GE Biosciences) column that was previously equilibrated with 20 mM Tris-HCl (pH 8.0), 1 M NaCl, and 30 mM imidazole. The column was extensively washed with the same buffer and elution of specifically bound protein was carried out using a gradient of increasing imidazole concentration. Fractions containing protein of the highest purity (as determined by SDS-PAGE) were pooled and the polyhistidine affinity tag was removed by overnight incubation with thrombin at 4 °C. Samples were further purified using size exclusion chromatography (Superdex Hiload 200 16/60) in a buffer of 20 mM HEPES-KOH, pH 7.5, 300 mM KCl. Samples were concentrated using a 10,000 Da molecular weight cutoff Amicon centrifugal filters. Selenomethionine-labeled (SeMet) MmFAMT was produced by repression of methionine biosynthesis in defined media prior to protein induction [28], and was purified as described above. All proteins were flash frozen in liquid nitrogen and stored at -80 °C. Prior to freezing, 2.5% glycerol (final concentration) was added to samples of wild type and mutant proteins used for kinetic analysis, but not to samples used for crystallographic studies.

Crystallization

Initial crystallization conditions were determined by the sparse matrix sampling method using commercial screens. Crystals of the MmFAMT-SAH complex were grown using the hanging vapor drop diffusion method. Briefly, 0.9 µl of protein at 4 mg/ml concentration was incubated with 2 mM SAM for 1 h on ice, and was subsequently mixed with 0.9 µl of precipitant solution (0.1 M $MgCl_2$, 0.1 M ADA, pH 6.5, 12% PEG 6,000) and 0.2 µl of 7% v/v 1-butanol, and equilibrated over a well containing the same precipitant solution at room temperature. Crystals grew within 2 days, and were transiently soaked in precipitant solution supplemented with 30% glycerol prior to vitrification by direct immersion in liquid nitrogen. Ligand bound crystals were

grown similarly using protein that was incubated with 2mM SAH, 2mM (3-hydroxy) fatty acid and 1mM DTT for 30 minutes prior to crystallization.

Data Collection, Phasing and Structure Determination

X-ray diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT), Sector 21, Argonne National Laboratory. All data were indexed, integrated and scaled using either HKL2000 or XDS. Initial crystallographic phases were determined by single wavelength anomalous diffraction from the SeMet-labeled protein crystals of the MmFAMT-SAH complex using data collected near the Se absorption edge (λ=0.97872 Å). Due to the low symmetry of the SeMet crystals (monoclinic setting) care was taken during crystal alignment to maximize the simultaneous collection of Bijvoet pairs. A four-fold redundant data set was collected to 1.95 Å resolution (R-merge of 5.5%, I/σ(I) of 2.2 in the highest resolution shell) using a MAR CCD detector. The heavy atom substructure was determined using SHELX [29]. The heavy atom positions were imported into SHARP [30] for maximum likelihood refinement resulting in preliminary phases with a mean Figure of Merit of 0.382. Density modification using solvent flattening and non-crystallographic symmetry averaging yielded a map of exceptional quality, which allowed for near all of the main chain and roughly half of the side chains to be automatically built using Buccanneer [31, 32] as implemented in the CCP4 suite of programs [33]. Additional cycles of manual rebuilding interspersed with crystallographic refinement using REFMAC5 [34] resulted in the final model. The final cycles of model building and refinement were carried out against high-resolution data collected on native crystals of MmFAMT-SAH.

Cocrystal structures with bound ligands were all determined by molecular replacement as implemented in the Phenix program suite [35], using the refined coordinates of SeMet MmFAMT

as search model. The resultant solutions were subsequently used as starting models for several rounds of automated model building using the ARP/wARP web server [33, 34, 36], followed by rounds of manual rebuilding using Coot [37], combined with crystallographic refinement using REFMAC5 [34]. Ligands were built in Coot, and water molecules were added using the ARP/wARP solvent building software of the CCP4 suite [38], and confirmed by manual inspection. In all cases, the quality of the in-progress model was routinely monitored using both the free R factor [39] and Molprobity [40] for quality assurance.

Determination of Kinetic Parameters

For all experiments, freshly purified protein was used, and SAM was further purified by HPLC using a C18 column (Vydac; 5 μm particle size, 4.6 × 250 mm) and monitoring absorbance at 260nm. The column was washed for 30 minutes with solvent B (methanol with 0.1% trifluoroacetic acid), and equilibrated for 15 minutes with solvent A (water with 0.1% trifluoroacetic acid). SAM was injected into the column and a gradient elution was applied as follows: wash with 5 ml of solvent A, elute with a linear gradient to a final 20% of solvent B, and wash with 5 ml of solvent B. The flow rate was 1 ml/min throughout the procedure. The fraction of SAM collected was subsequently lyophilized, and stored in -20 °C. Fresh solution of SAM was prepared before each experiment.

The kinetic parameters of the wild type and mutant proteins were determined by using a photospectrometric assay that monitors the production of SAH [41, 42]. All enzyme reactions were performed in 100 mM HEPES-KOH (pH 7.8) and 300 mM KCl at 37 °C, and monitored at 265 nm for up to 20 min. Control reactions without addition of the substrate were also included to take into account any background hydrolysis of SAM over time. For determination of the Michaelis-

Menten parameters of the wild type enzyme and mutants for the fatty acids and 3-hydroxy fatty acids, a 150 µl reaction volume contained the following components: 0.1 – 2.0 µM MmFAMT, 1 µM SAH nucleosidase, 0.2 µM adenine deaminase, 1 mM $MnSO_4$, 80 µM SAM, and various concentrations of fatty acids and 3-hydroxy fatty acids. For determination of the kinetic parameters for SAM, a 150 µl reaction volume contained 0.2 µM MmFAMT, 1 µM SAH nucleosidase, 0.2 µM adenine deaminase, 1 mM $MnSO_4$, 3 mM of octanoic acid (C8) and various concentrations of SAM. Based on the initial reaction rates, the apparent K_M and V_{max} values were determined using the Michaelis-Menten function of Origin (OriginLab, Northampton, MA). Results are means ± SEM of triplicate experiments.

For mutants with very increased K_M values for C8 or 3-hydroxy-decanoic acid (3-OH-C10), saturation of the enzyme could not be achieved due to limited solubility of the substrates in the reaction buffer. In these cases, the apparent K_M and V_{max} values could not be determined, but the k_{cat}/K_M values were obtained by plotting the observed rates (k_{obs}) at 4 different substrate concentrations (Figure 1. 2A). For the Tyr24→Ala mutant, the kinetic parameters for SAM, and consequently for C8 and 3-OH-C10, could not be determined as the K_M value increased such that saturation of the enzyme could not be achieved due to limitations of the assay (concentration of SAM used should be kept below 250 µM to remain in the linear range of the spectrophotometer [41]).

To attest that the coupling enzymes were not rate-limiting, the initial rates of 0.1, 0.2 and 0.4 µM of wild type enzyme were determined by addition of 3 mM C8. The means of the observed rates ± SEM were found to be the same (Figure 1. 2A), indicating that the coupling enzymes are not rate limiting. Consequently, the measured rate corresponds to the rate of MmFAMT.

All substrates were purchased from Sigma-Aldrich, except for the 3-OH-C8, which was purchased from Matreya, LLC. The 3-hydroxy fatty acids were purchased as enantiomeric mixtures. However, it is clear from the crystal structure of the enzyme complexed with SAH and 3-OH-C10 that the enzyme utilizes only the S enantiomer (Figure 1. 6C and Figure A. 2). Subsequently, the enantiomeric ratio was determined by Mosher ester analysis [43], and was found to be 1:1.

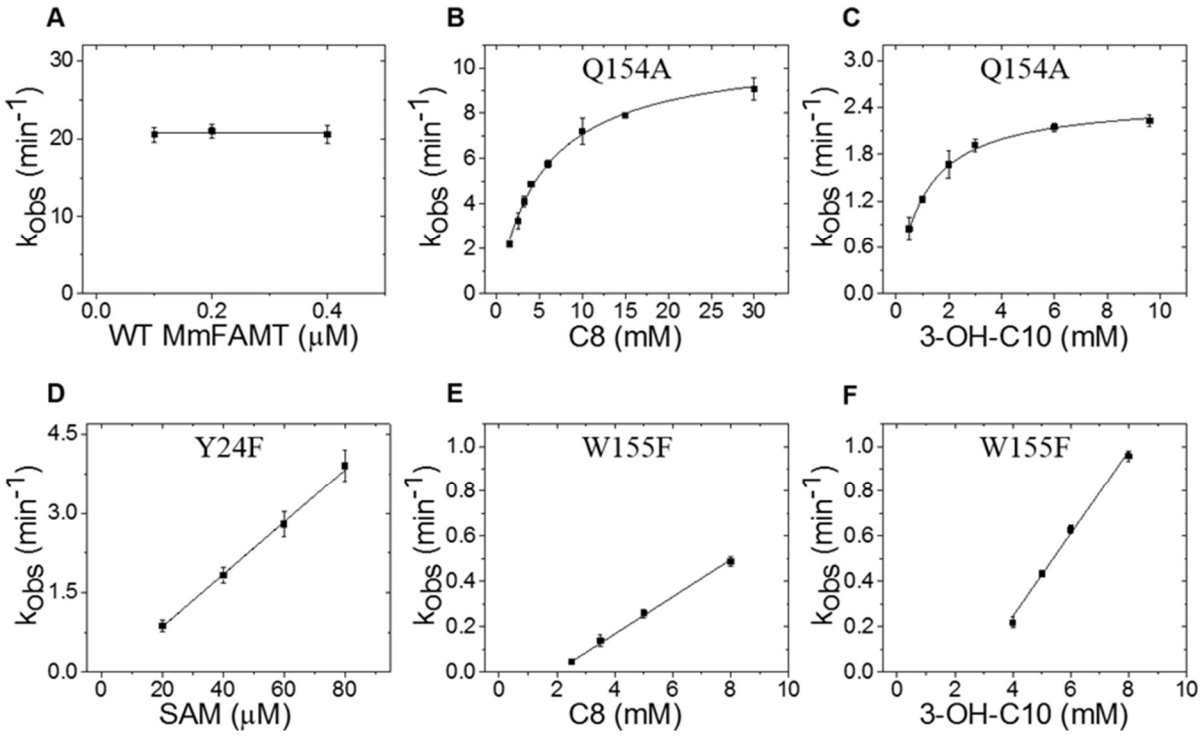

Figure 1. 2: (A) Observed rates at 3 different enzyme concentrations for 3mM octanoate. All rates are the same indicating that the coupling enzymes are not limiting. (B, C) Michaelis-Menten kinetic plots for the mutant Q154A for octanoate (C8) and 3-hydroxydecanoate (3-OH-C10). (D, E, F) Observed rates at 4 different enzyme concentrations for the Y24F and W155F mutants. The slope of the plot corresponds to the k_{cat}/K_M value.

End-Point Activity Assay

The activity of the wild type enzyme for salicylic acid (SA) was investigated by gas chromatography-mass spectroscopy (GC/MS). 100 µl reactions containing 50 µM wild type MmFAMT, 500 µM SAH nucleosidase, 5 mM SAM and 10 mM substrate were incubated at 37 °C for 2 h. The samples were cleaned from the enzyme with a 10,000 Da molecular weight cutoff Amicon spin filter, and analyzed by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign). All reactions were performed in the same buffer used for kinetic analysis supplemented with 5% methanol (final concentration). To confirm that methanol did not deactivate the enzyme and possible observed inactivity was not due to its addition, reactions containing wild type MmFAMT and C8 were included. Control reactions without addition of the enzyme were also analyzed. Results are means ± SEM of triplicate experiments.

Isothermal Titration Calorimetry (ITC)

The binding affinity of wild type MmFAMT for SAM and SAH was measured at 25 °C using a VP-ITC microcalorimeter (Microcal Inc., Northampton, MA). Protein and ligands were in the same buffer used for kinetic analysis. For binding of SAM, 1.1 – 1.15 mM SAM was injected into the reaction cell containing 30 – 35 µM protein in 28 successive aliquots at 300 s intervals and 20.5 s duration with a reference power of 2 µCal/s. For binding of SAH, 0.52 – 0.54 mM SAH was injected into the reaction cell containing 40 – 42 µM protein in 28 successive aliquots at 240 s intervals and 20.5 s duration with a reference power of 6 µCal/s. All injections were 10 µl in volume, except for the first injection which was 4 µl and was excluded from data analysis. The protein-ligand buffer was used in the reference cell, and a titration of the ligand into just the buffer was subtracted from the measurements. Non-linear regression with a single-site fitting model

(MicroCal Origin) was applied for data analysis, and the thermodynamic parameters were calculated using the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$) and the relationship $\Delta G = -RT\ln K_a$. Results are means ± SEM of duplicate experiments.

For the binding experiments, freshly purified protein was used, and SAM was further purified by HPLC using a C18 column (Vydac; 5 µm particle size, 4.6 × 250 mm) and monitoring absorbance at 260 nm. The column was washed for 30 minutes with solvent B (methanol with 0.1% trifluoroacetic acid), and equilibrated for 15 minutes with solvent A (water with 0.1% trifluoroacetic acid). SAM was injected into the column and a gradient elution was applied as follows: wash with 5 ml of solvent A, elute with a linear gradient to a final 20% of solvent B, and wash with 5 ml of solvent B. The flow rate was 1 ml/min throughout the procedure. The fraction of SAM collected was subsequently lyophilized, and stored in -20 °C. Fresh solution of SAM was prepared before each experiment.

Sequence Similarity Network

A sequence similarity network was generated by using the Enzyme Function Initiative - Enzyme Similarity Tool (http://efi.igb.illinois.edu/efi-est/) [44] with the sequence of MmFAMT as the query for a BLASTP [45, 46] search of the UniProtKB database (http://www.uniprot.org/) [47]. Only sequences of 300-450 amino acids were included for the subsequent generation of a network with E-values equal to or lower than $1x10^{-30}$. This network was visualized using Cytoscape 3.2.1 (Figure 1. 8) [48].

1.3 Results and Discussion

Kinetic Characterization of *M. marinum* FAMT

The kinetic parameters for wild type *M. marinum* FAMT (MmFAMT) for a panel of fatty acid and 3-hydroxy fatty acids were determined using a coupled assay [41, 42] that monitored SAM turnover. In this assay, S-adenosylhomocysteine (SAH), the product of SAM, gets hydrolyzed by an SAH nucleosidase (the SAH nucleosidase utilized has a catalytic efficiency of 11.6×10^6 $M^{-1}s^{-1}$ [49], which is greater than that of MmFAMT) to S-ribosylhomocysteine and adenine. The latter is subsequently deaminated by an adenine deaminase resulting in a decrease in absorbance at 265 nm ($\Delta\varepsilon \approx 6{,}700$ $M^{-1}cm^{-1}$), and allowing for continuous monitoring of the reaction [41]. The catalytic efficiency of MmFAMT for free fatty acids was found to increase with chain length (the k_{cat}/K_M values for C8 octanoic acid is 1.49×10^3 $M^{-1}s^{-1}$; and for C10 decanoic acid is 2.97×10^4 $M^{-1}s^{-1}$) (Table 1. 1 and Figure 1. 3). The 3-hydroxy compounds were turned over less efficiently, with k_{cat}/K_M values that are roughly an order of magnitude lower than for the corresponding fatty acid. The kinetic parameters for MmFAMT reported here differ from values determined previously via a discontinuous assay that employed radiolabeled SAM as a donor [9]. These differences likely arise from variations in the two methods used for characterization. However, in the context of this work, our kinetic data provide an internal standard for analyzing the effects of specific amino acid mutations on the catalytic activity of MmFAMT as reported below.

Table 1. 1: Steady state kinetic parameters for MmFAMT wild type and mutants against different substrates.

MmFAMT	K_M (μM)	k_{cat} ($10^{-2}s^{-1}$)	k_{cat} / K_M ($M^{-1}s^{-1}$)
WT			
C8	308 ± 9	46.1 ± 0.5	1,497
C10	8.2 ± 0.5	24.4 ± 0.4	29,756
3-OH-C8	2,039 ± 172	48.8 ± 1.4	239
3-OH-C10	201 ± 8	59.9 ± 1.1	2,975
SAM	25.5 ± 4.0	59.8 ± 3.2	23,450
Q31A			
C8	*	*	*
3-OH-C10	*	*	*
Q154A			
C8	5,238 ± 427	14.7 ± 0.4	28.1
3-OH-C10	1,028 ± 43	4.2 ± 0.0	40.9
W155F			
C8	*	*	1.37 ± 0.03
3-OH-C10	*	*	3.05 ± 0.16
Y24F			
SAM	*	*	824 ± 19

*** Could not be determined**

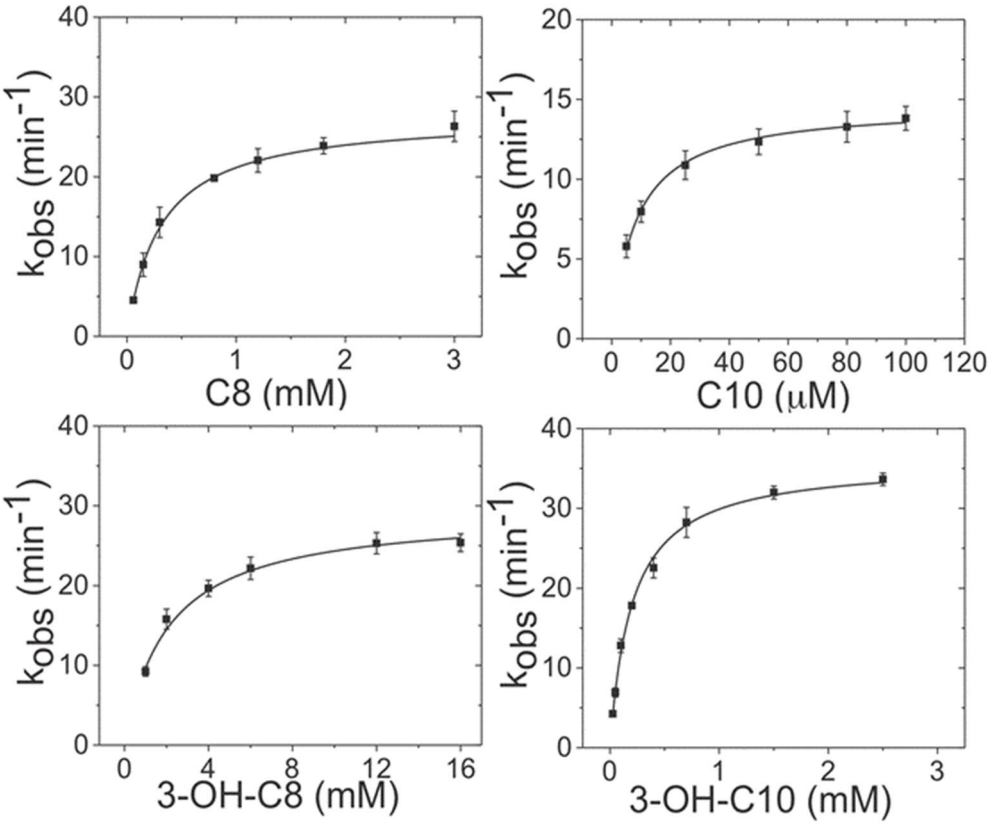

Figure 1. 3: Michaelis-Menten kinetic plots for the wild type MmFAMT for octanoate (C8), decanoate (C10), 3-hydroxyoctanoate (3-OH-C8) and 3-hydroxydecanoate (3-OH-C10).

Binding Affinity of MmFAMT for SAM and SAH

The product SAH is known to inhibit some SAM-dependent methyltransferases. In order to characterize if MmFAMT is similarly subject to product inhibition, we carried out isothermal titration calorimetric analysis to measure the binding affinity to SAM and SAH (Table 1. 2 and Figure 1. 4). MmFAMT binds to SAM with a dissociation constant (K_d) of 33.6 μM, while binding of SAH occurs with a K_d that is two orders of magnitude smaller (0.70 μM). For both ligands, binding is driven largely by the enthalpic component. These data are consistent with the inhibition of MmFAMT by the product SAH. It should be noted that all measurements using SAM as the ligand yielded a number of binding sites of (N = 2.06 +/- 0.01), while measurments with SAH gave a number of sites of (N = 0.93 +/- 0.01). The basis for this discrepancy is not clear, as the structural

data do not reveal evidence for multiple binding sites. Nonetheless, it is evident from the binding isotherms that MmFAMT binds tightly to SAH. These results are in agreement with product inhibition observed in several members of this family [50-52]. Hence, both *in vivo* and *in vitro* efficacy of enzymatic methylation of free fatty acids using FAMTs may be improved if the reactions were coupled to an S-adenosylhomocysteine nucleosidase to hydrolyze the SAH product.

Table 1. 2: ITC binding parameters for wild type MmFAMT to SAH and SAM.

Ligand	N	K_a (M^{-1}) x10^4	K_d (M) x10^{-7}	ΔH (cal/mol) x10^3	ΔS (cal/mol/deg)
SAH	0.93 ± 0.01	143.5 ± 3.5	6.97 ± 0.17	-22.3 ± 0.6	-46.5 ± 2.1
SAM	2.06 ± 0.01	3.0 ± 0.4	335.8 ± 45.0	-4.4 ± 0.3	5.8 ± 0.6

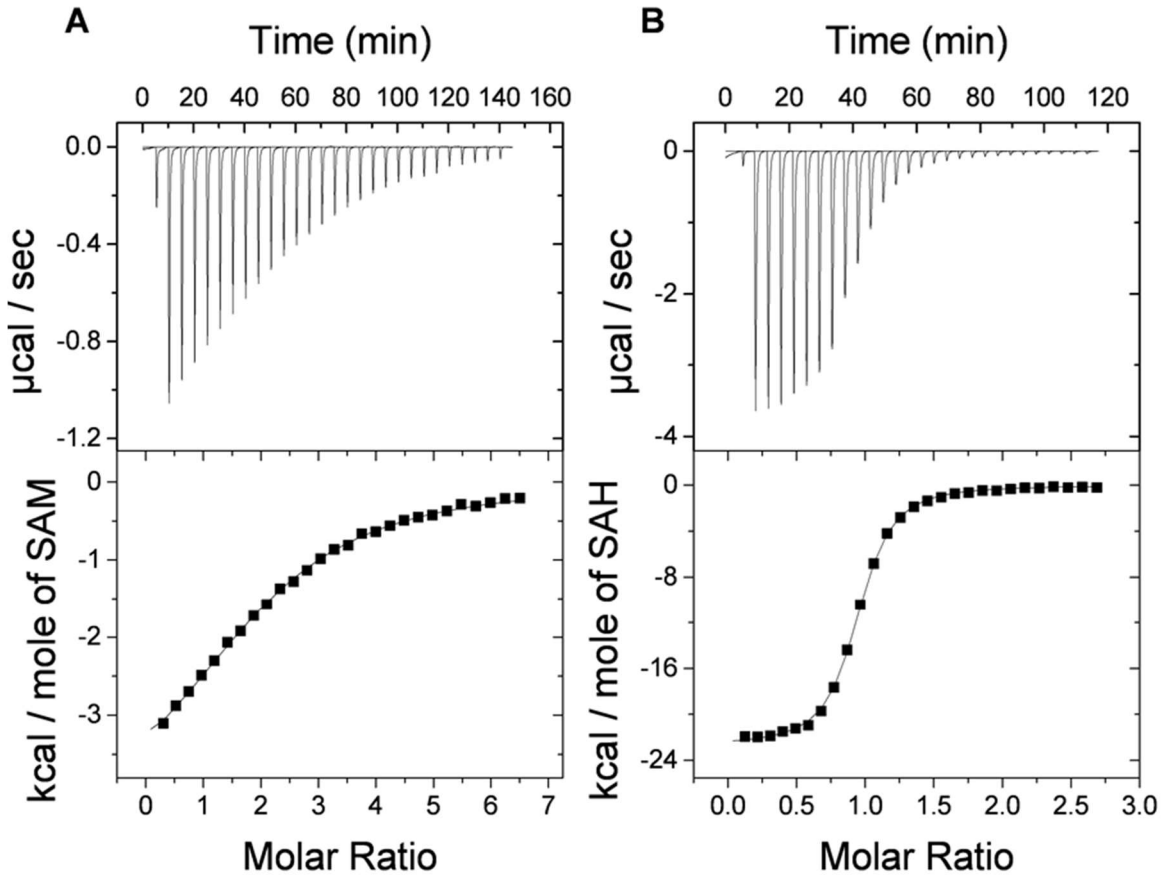

Figure 1. 4: Representative ITC data for binding of MmFAMT to (A) SAM and (B) SAH.

Crystal Structures of MmFAMT

The unique specificity of bacterial FAMTs for fatty acid substrates cannot be understood in the context of the distantly related plant SABATH family of methyltransferases, which share limited sequence similarities (25% sequence identity) and function on small molecule natural products [24]. In order to understand this substrate specificity and scope, we determined several binary and ternary complex structures of MmFAMT in complex with SAH (1.7 Å resolution), SAH and octanoate (1.6 Å resolution), and SAH and 3-hydroxydecanoate (1.9 Å resolution). We also determined the structure of an MmFAMT active site mutant (Gln154→Ala) in complex with SAH and 3-hydroxydecanoate (1.85 Å resolution). Initial crystallographic phases were determined by single-wavelength anomalous diffraction using SeMet labeled protein (SAH and decanoic acid), and phases for subsequent structures were determined by molecular replacement using the resultant model as a search probe. All crystallographic data collection and refinement statistics are provided in Table 1. 3.

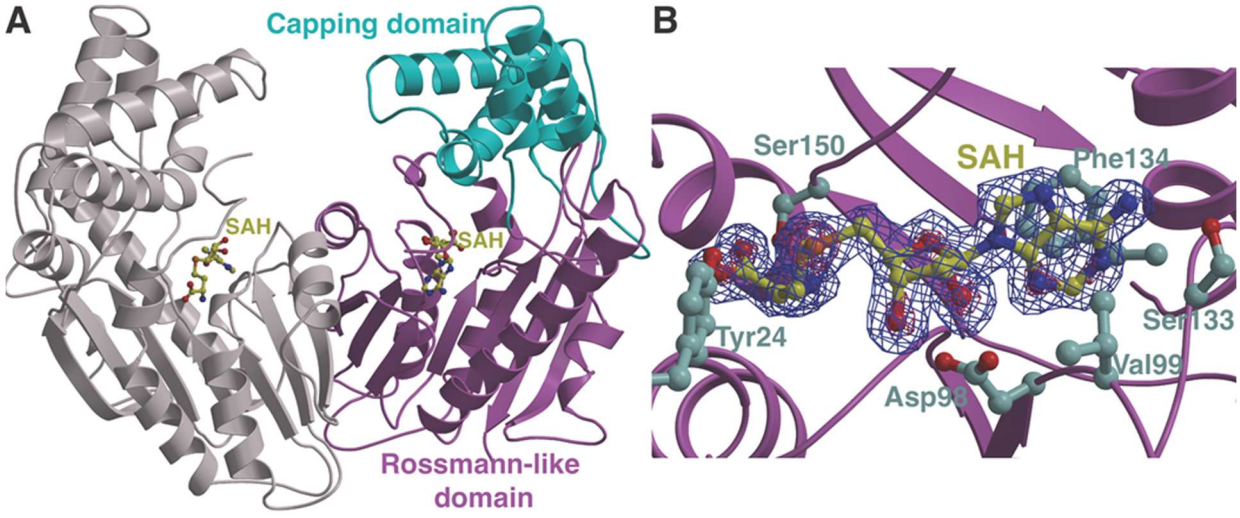

Figure 1. 5: (A) Ribbon diagram of the MmFAMT-SAH crystal structure with one monomer colored in gray and the other colored in purple (Rossmann fold-like domain) and cyan (capping domain). (B) Simulated annealing difference Fourier map (Fo-Fc) contoured to 2.5σ (blue) and 10σ (red) showing the SAM-binding site of MmFAMT. The coordinates for bound SAH were omitted during map calculations. The bound ligand is colored in yellow and residues that interact with the ligand are colored in gray.

Table 1. 3: Data collection, phasing and refinement statistics (1. Highest resolution shell is shown in parenthesis, 2. $R_{sym} = \Sigma |(I_i - \langle I_i \rangle| \Sigma I_i$ where I_i = intensity of the *i*th reflection and $\langle I_i \rangle$ = mean intensity, 3. Mean figure of merit (acentric/centric), 4. R-factor = $\Sigma (|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement).

	Native	**SeMet**	**C8**	**3-OH-C10**	**Q154A-3-OH-C10**
Data collection					
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å), β (°)	63.1, 66.1, 98.4, 107.5	62.7, 65.7, 97.5, 107.8	63.1, 66.1, 98.1, 107.4	62.8, 65.9, 98.2, 107.1	62.8, 65.6, 98.2, 107.3
Resolution (Å)[1]	50-1.7 (1.73-1.7)	50-1.95 (1.98-1.95)	50-1.6 (1.63-1.6)	50-1.9 (1.93-1.9)	50-1.85 (1.88-1.85)
R_{sym} (%)[2]	5.4 (52.1)	5.5 (58.8)	6.0 (57.3)	6.5 (54.1)	7.0 (81.4)
I/σ(I)	25.6 (2.6)	15.9 (2.2)	20.5 (1.9)	18.2 (2.1)	14.5 (1.9)
Completeness (%)	100 (99.9)	100 (100)	100 (99.7)	99.5 (100)	100 (100)
Redundancy	5.0 (4.8)	4.1 (4.1)	4.2 (3.5)	4.1 (4.1)	4.7 (4.7)
Phasing					
FOM[3]		0.382/0.157			
Refinement					
Resolution (Å)	25.0-1.7		25.0-1.6	25.0-1.9	25.0-1.85
No. reflections	80,265		95,148	55,801	61,413
R_{work} / R_{free}[4]	19.5/22.2		19.5/21.8	19.3/23.1	19.3/22.4
Number of atoms					
Protein	5504		5533	5525	5525
Fatty acid	-		20	26	44
SAH	52		52	52	52
Water	793		802	563	570
B-factors					
Protein	16.7		16.9	24.1	23.9
Fatty acid	-		16.2	17.2	22.2
SAH	7.9		8.1	12.8	12.6
Water	27.9		28.4	30.9	31.2
R.m.s deviations					
Bond lengths (Å)	0.000		0.005	0.006	0.006
Bond angles (°)	1.07		1.07	1.15	1.11

The overall structure of MmFAMT consists of a di-domain architecture composed of a Rossmann-like α/β fold that is common amongst diverse class I methyltransferases, which is

further elaborated with an all α-helical domain that caps the SAM-binding site (Figure 1. 5A). The Rossmann-like fold core is formed from discontinuous regions of the polypeptide chain and space residues Ser11 through Asp217, Pro255 through Ala287, and Pro354 through the carboxy-terminus. This core consists of a seven-stranded β-sheet core that is sandwiched between sets of helical regions. The helical capping domain is similarly composed of disjointed segments and forms part of the active site, in a fashion similar to that found in plant natural product methyltransferases [25, 26, 53].

Despite a low conservation in primary sequence, a search of the Protein Data Bank [54] using the Dali server [55] shows that MmFAMT is structurally homologous to the SABATH class of plant natural product methyltransferases, which themselves are only distantly related to other well-characterized small molecule methyltransferases [27]. Representative structural homologs include the salicylic acid *O*-methyltransferase from *Clarkia breweri* (PDB code 1M6E; RMSD of 2.5 Å over 330 Cα atoms; 23% sequence identity) [26], the dimethylxanthine N-methyltransferase from *Coffea canephora* (PDB code 2EFJ; RMSD of 3.2 Å over 328 Cα atoms; 22% sequence identity) [25], and indole-3-acetic acid *O*-methyltransferase from *Arabidopsis thaliana* (PDB code 3B5I; RMSD of 2.6 Å over 312 Cα atoms; 23% sequence identity) [53] (Figure A. 3). The unexpected structural similarities between MmFAMT and the plant SABATH *O*-methyltransferases reflect the similar substrate repertoires of these enzymes, specifically that they both catalyze methylation of carboxylic acid on an otherwise hydrophobic substrate. However, the nature of the hydrophobic groups differs as the plant enzymes utilize larger, bulkier scaffolds that are typically aromatic, while the FAMTs function on fatty acids containing a long hydrocarbon acyl chain (Figure 1. 1).

The plant SABATH methyltransferases described above are all homodimers in solution and the corresponding crystal structures demonstrate a two-fold symmetric arrangement, and MmFAMT likewise is homodimeric both in solution and in the crystal lattice (Figure A. 1). In plant *O*-methyltransferases that are not members of the SABATH class, notably chalcone *O*-methyltransferase, and caffeic acid *O*-methyltransferase, the dimeric arrangement is functional with residues from one monomer contributing to the active site of the other subunit [56, 57]. Although the SABATH enzymes are also dimeric, oligomerization is not needed for activity. Oligomerization of MmFAMT is similarly likely not functional as the monomeric entity contains all of the necessary residues involved in catalysis. Likewise, dimerization only buries ~1,314 $Å^2$ of solvent-accessible surface area (corresponding to 8% of the total surface area of each monomer), which is far less than for the functionally homodimeric enzymes. The dimerization interface of MmFAMT is similar in organization to that of the salicylic acid *O*-methyltransferase [26] and indole-3-acetic acid *O*-methyltransferase [53], both of which also likely function as a monomer.

SAM/SAH and Substrate Binding Pockets

Like in other class I methyltransferases, the SAM/SAH binding site is situated in the Rossmann-like fold domain, and the ligand is bound in an extended manner, roughly perpendicular to the plane of the β-strands. In the MmFAMT-SAH structure, the pyrimidine ring of SAH adenine is sandwiched between the side chain of Phe134, via a π-stacking interaction, on one side and Val99 on the other (Figure 1. 5B). Extensive hydrogen-bonding interactions further stabilize the bound SAH, including the interaction between Ser133 and the adenine amine, between Asp98 and both hydroxyls of the ribose, and between Tyr24 and Ser150 and the α-carboxylate of SAH. Additional contacts are mediated through van der Waals interactions with non-polar residues that

contribute to form the SAM/SAH binding site. An analogous network of interacting residues stabilizes bound SAM/SAH in structures of the plant natural product *O*-methyltransferases. The extensive set of interactions and the hydrophobicity of the binding site rationalize why uncharged SAH is a strong competitive inhibitor of MmFAMT.

The cocrystal structure of MmFAMT with SAH and octanoate shows that the substrate-binding site is largely localized within the α-helical domain that caps the Rossmann-like fold domain (Figure A. 3). This capping domain contains a binding cavity that is sufficiently contoured to accommodate fatty acid substrates with acyl chain lengths up to C12. Numerous aliphatic and aromatic residues encircle the acyl chain of the fatty acid substrate and these include Trp155, Met214, Phe222, Tyr224, Val256, Phe311, and Leu316 (Figure 1. 6A). MmFAMT residues Tyr174, Met 227 and Asn228 enclose the upper side of the binding cavity, and multiple conformers are observed for the latter two residues. The planar, extended conformation of the fatty acid substrate is set by the side chains of Met19, Ala315, and Leu316 on one side of the substrate and by Trp151, and Tyr224 on the opposite side of the substrate. In the salicylic acid *O*-methyltransferase cocrystal structure, Met150 and Met308 similarly sandwich the substrate for proper positioning at the SAM methyl donor [26]. Lastly, hydrogen bonding with Gln31 and Trp155 engage and orient the carboxylate moiety of the fatty acid substrate for methyl transfer, similar to the Gln25/Trp151 pair utilized by salicylic acid *O*-methyltransferase to orient the carboxylate of its substrate. The relative disposition of the α-helical capping and Rossmann-fold domains is slightly different than that observed in structures of plant *O*-methyltransferases with a similar bi-lobed structure, which may also contribute to different substrate scope of the FAMTs (Figure A. 3).

A comparison of the MmFAMT/SAH cocrystal structures with bound octanoate and 3-hydroxydecanoate suggests a rationale for understanding why MmFAMT shows a preference for the larger substrate (K_M values of C10 substrates lower by a factor of ~10-40 relative to the C8 counterparts; Table 1. 1). While Met227 and Asn228 are observed as multiple conformers in the cocrystal structure with octanoate (Figure 1. 6B), both of these residues exist in single conformations oriented away from the substrate-binding cavity to accommodate the larger acyl chain in the SAH/3-hydroxydecanoate structure (Figure 1. 6D). This movement is accompanied by a compensatory reorganization of a loop spanning residues Leu166 through Gln171, located near the omega carbon of the fatty acid, resulting in tighter packing against C9 and C10 of the acyl chain (Figure 1. 6Figure 1.6D). Movement of these side chains fills the substrate-binding cavity but only with acyl chains of length C10-C12, and likely explains why these longer chain fatty acids are better substrates than octanoate (Figure 1. 6B and Figure 1. 6D). There are limited interactions in the active site with the 3-hydroxyl moiety, as Gln154 is located within hydrogen bonding distance and the thioether of Met214 is over 3.2 Å away. Placement of the polar 3-hydroxyalcohol in a hydrophobic binding pocket without sufficient compensatory interactions may explain the higher K_M values for 3-hydroxy-containing substrates, relative to their acyl counterparts (Table 1. 1). Lastly, Gln154 is positioned to hydrogen bond with the carboxylate of the substrate in the MmFAMT/SAH/octanoate structure. However, a modest rotation of the plane of the substrate, necessary to maximize interactions of the 3-hydroxyl group with Gln154, results in a slight movement of this residue away from the substrate carboxylate in the MmFAMT/SAH/3-hydrodecanoate structure.

In the cocrystal structure of MmFAMT with 3-hydroxydecanoate, only the S enantiomer was bound in the active site although an enantiomeric mixture of 50% R and 50% S was used in the crystallization condition. This enantiomeric selectivity is due to the presence of hydrophobic residues (Phe311 and Trp151) on one side of the active site in close proximity to the carboxyl

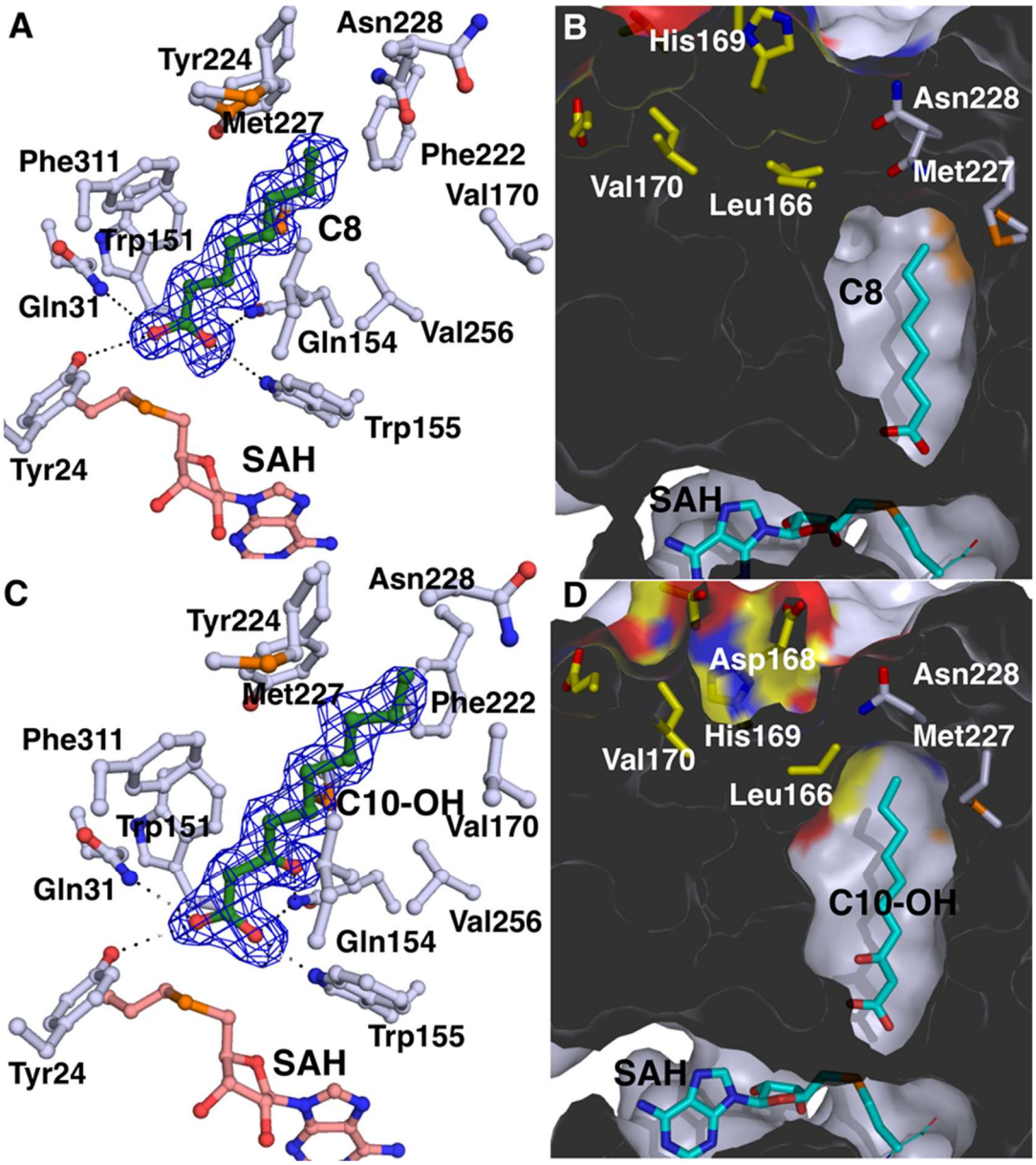

Figure 1. 6: (A, C) Difference Fourier maps (Fo-Fc) contoured to 2.5σ (blue) showing the bound fatty acid substrates (colored in green). (B, D) Surface cut-away diagram showing that binding of the larger substrate results in a smaller cavity due to movement of Leu166 through Gln171.

group, and the presence of Gln154 on the other side in a position favorable for hydrogen bonding with the 3-hydroxyl group of the S enantiomer. Although the physiological role of MmFAMT is yet to be elucidated, this enantiomeric specificity may be of physiological relevance. 3-hydroxylated fatty acids are found in mycobacterial phospholipids [58, 59], enter β-oxidation as (S)-3-hydroxy-acyl-Coenzyme A [60, 61], and are part of the mycolic acid biosynthesis as (R)-3-hydroxy-acyl-ACP (Acyl Carrier Protein) [62, 63]. Methylation of the carboxyl group of the (S)-3-hydroxy fatty acids may be part of any of these processes by regulating the fate of the (S)-3-hydroxy-fatty acids. This hypothesis, however, is to be investigated.

Molecular Basis for Fatty Acid Substrate Specificity

As noted, the structure of MmFAMT shares several features with that of salicylic acid *O*-methyltransferase [26], including the presence of a hydrophobic α-helical domain that caps the SAM-binding site. While both enzymes utilize a similar constellation of active site residues to orient the substrate carboxylate (Gln31/Trp155 in MmFAMT and Gln25/Trp151 in salicylic acid *O*-methyltransferase), the capping domain establishes specificity for the hydrophobic part of the substrate (Figure 1. 7A and Figure 1. 7B). Specifically, the contours of each cavity are optimized for its cognate substrate. In salicylic acid *O*-methyltransferase the active site is border lined by Ile 225, Trp226, Tyr255 and Phe347, while in MmFAMT by Tyr174, Phe222, Tyr224, Met227 and Asn228. The replacement of Ile225 and Trp226 (in salicylic acid *O*-methyltransferase) by Met227 and Asn228 (in MmFAMT) creates a larger cavity to accommodate the longer acyl chain of the fatty acid (Figure 1. 6, Figure 1. 7A, and Figure 1. 77B).

Although the constrained smaller cavity of salicylic acid *O*-methyltransferase may not accommodate a fatty acid, the larger cavity of MmFAMT raised the question as to whether

MmFAMT can utilize salicylic acid (SA) as a substrate despite the lack of residues that would anchor the aromatic ring in an optimal orientation. In order to further investigate this, we carried out reactions (2 h incubation at 37 °C) of MmFAMT with SA in triplicate, and analyzed the

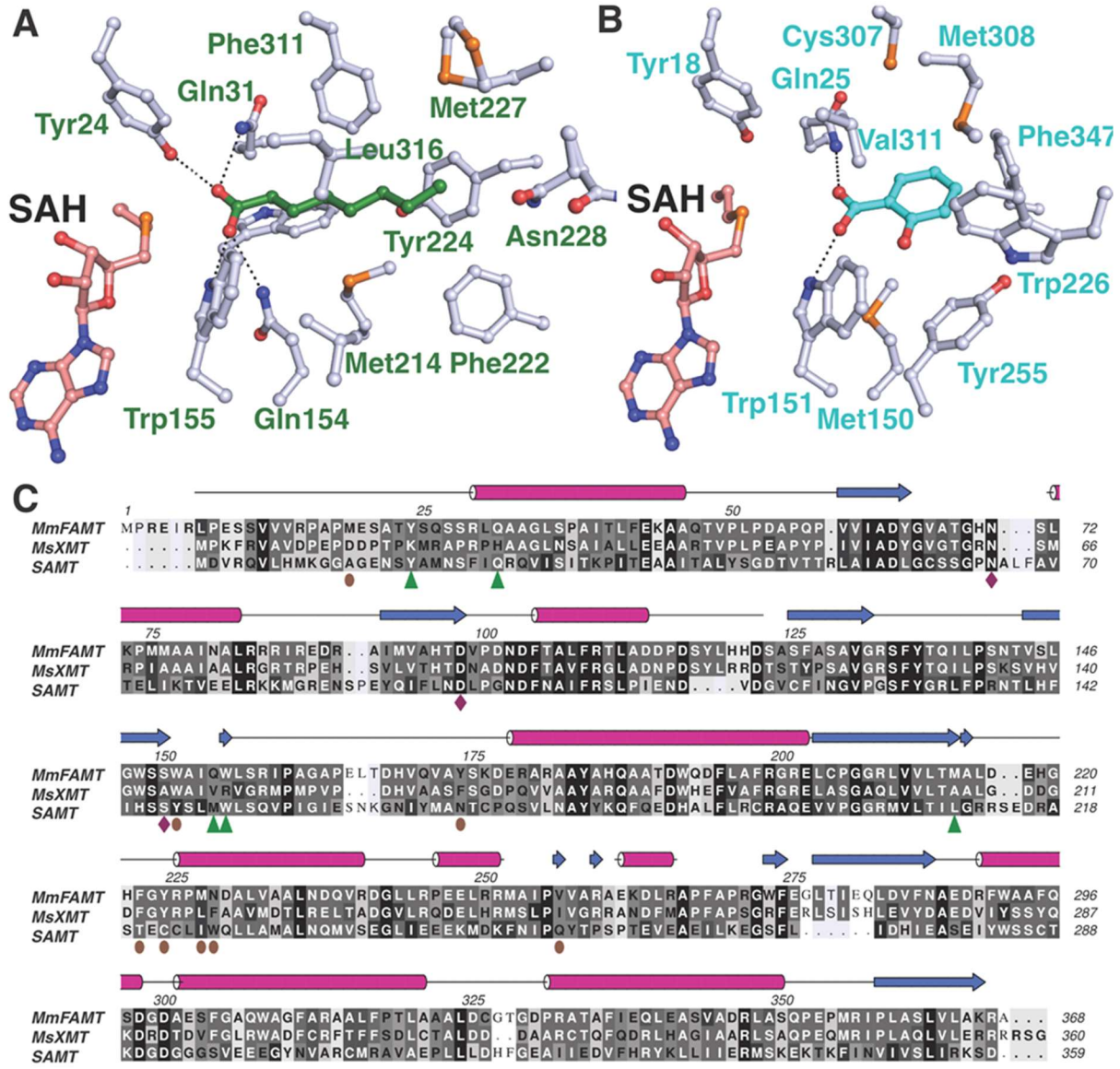

Figure 1. 7: Comparison of the substrate binding active site in (A) MmFAMT and (B) the SABATH family enzyme salicylic acid *O*-methyltransferase. (C) Structure based alignment of the primary sequences of MmFAMT, MsXMT and salicylic acid O-methyltransferase. The green triangles demarcate residues involved in interactions with the carboxylic acid, the brown circles demarcate residues that form the fatty acid binding pocket, and the magenta diamonds indicate residues that interact with SAH/SAM.

products by gas chromatography-mass spectrometry (GC/MS). Surprisingly, MmFAMT is able to catalyze methylation on SA (Table A. 2). A closer examination of an active site superposition of the cocrystal structures of MmFAMT-SAH-octanoate and salicylic acid *O*-methyltransferase-SAH-SA reveals that SA can be accommodated into the MmFAMT active site. Although the positioning of Phe311 (Cys307 in salicylic acid *O*-methyltransferase) would seemingly cause some steric clashes, the residue at the other side of the aromatic ring (Tyr255 in salicylic acid *O*-methyltransferase) is a smaller Val256 in MmFAMT, which would allow for positioning of SA as a substrate. Equivalent, but distinct, residues in the MmFAMT active site could orient SA. However, in salicylic acid *O*-methyltransferase, Ile225 and Trp226, which are replaced by the aforementioned flexible Met 227-Asn228 in MmFAMT, constrict the upper end of the active site to accommodate the smaller SA substrate (Figure 1. 7A and Figure 1. 7B). Consequently, it is likely that SA may not bind very well in the MmFAMT active site.

Kinetic Analysis of Wild Type and Mutants MmFAMTs

The cocrystal structures of MmFAMT identified a number of residues that may play a role in either substrate binding or catalysis. In order to further probe function, we generated site-specific mutations at several of these residues, and determined kinetic parameters for the variant enzymes (Table 1. 1). The Gln31→Ala mutation had an immense effect on the enzyme activity as no product formation could be detected even when 20-fold more enzyme and much greater concentrations (up to 5 mM) of substrate were used. Similarly, the Trp155→Phe mutation resulted in a near 1000-fold decrease in catalytic efficiency (k_{cat}/K_M). As Gln31 and Trp155 are situated near the carboxylate of the substrate, the Gln31→Ala and Trp155→Phe mutations likely compromise binding and orientation of the substrate, consistent with the loss of activity in each of these mutants.

The Gln154→Ala mutation had a large effect on the K_M of MmFAMT with octanoate as a substrate (17-fold increase) and a smaller effect on the K_M for 3-hydroxydecanoate (5-fold increase). This was an unexpected result as Gln154 is within hydrogen bonding distance to the 3-hydoxyl group, and a mutation at this residue would be expected to cause a much greater increase in the K_M for the 3-hydroxy fatty acid. In order to provide a rationale for this observation, we determined the 1.85 Å resolution co-crystal structure of MmFAMT Gln154→Ala in complex with SAH and 3-hydroxydecanoate. The structure shows that the 3-hydroxyl group rotates towards the sulfur of Met214 to compensate for the loss of interaction with Gln154 (Figure A. 2), explaining why this mutation results in only a modest increase in the K_M for 3-hydroxydecanoate. The Gln154→Ala mutation also results in the carboxylate shifting away to result in a less-optimal orientation for methyl transfer, consistent with a 10-fold lower k_{cat} for the mutant relative to wild-type MmFAMT.

1.4 Concluding Remarks

Our collective structural and biochemical analysis demonstrates how MmFAMT uses a SABATH plant natural product methyltransferase architecture to catalyze the methylation of fatty acid substrates. This adaptation is a result of minor alterations in secondary structural elements, and of the changes in the disposition of the α-helical capping domain, which engages the substrate, relative to the Rossmann-like fold domain that harbors the methyl donor. Nonetheless, the identification of MmFAMT as a structural homolog of the SABATH class of methyltransferases extends the function of Pfam03492 [27] to beyond plant metabolism.

Prior studies on the plant SABATH members suggest that these enzymes do not require a general base to deprotonate the substrate methyl acceptor, as its carboxylate is likely ionized at

physiological pH due to its low pK_a [26]. Likewise, there are no residues in the active site of MmFAMT that can abstract the proton from the fatty acid carboxylate prior to methyl transfer, suggesting that the enzyme active site simply serves to facilitate proximity and orientation of the reactive groups. The side chain amide of Gln31 and the indole nitrogen of Trp155 are within hydrogen bonding distance from the two carboxylate oxygens, and structure-guided engineering studies [26] demonstrate that replacement of the equivalent Gln in plant *O*-methyltransferases establishes a selectivity against different methyl acceptors. The replacement of both of these residues in the MmFAMT homolog from *Mycobacterium smegmatis* (52% sequence identity, UniProt ID: A0R0D4) (MsXMT in Figure 1. 7C) may account for the lack of activity against fatty acid substrates in the latter [9].

Studies of plant SABATH enzymes have determined that small changes in the primary amino acid sequence of these enzymes can establish methyltransferase activity on structural scaffolds with diverse chemical structures. Similarly, small changes in the primary sequence of mycobacterial methyltransferases homologous to MmFAMT (Figure 1. 8 and Figure A. 4) provide a pool of enzymes that have been engineered by nature to utilize fatty acids of different chain lengths. The identification of the active site residues of MmFAMT that recognize the carboxyl of the substrate as well as the residues that line the hydrophobic pocket that harbors the fatty acid tail can guide the selection of MmFAMT homologs for the *in vivo* production of various fatty acid methyl esters of desired length.

Prior attempts to utilize MmFAMT for production of biodiesel in *E.coli* [9] resulted in yields that were considerably lower than other reported fatty acid ethyl esterification processes [7]. From the crystal structure of MmFAMT, as well as the previous *in vivo* studies [9], it is evident that MmFAMT can methylate medium chain fatty acids (up to 12-14 carbons long), which could

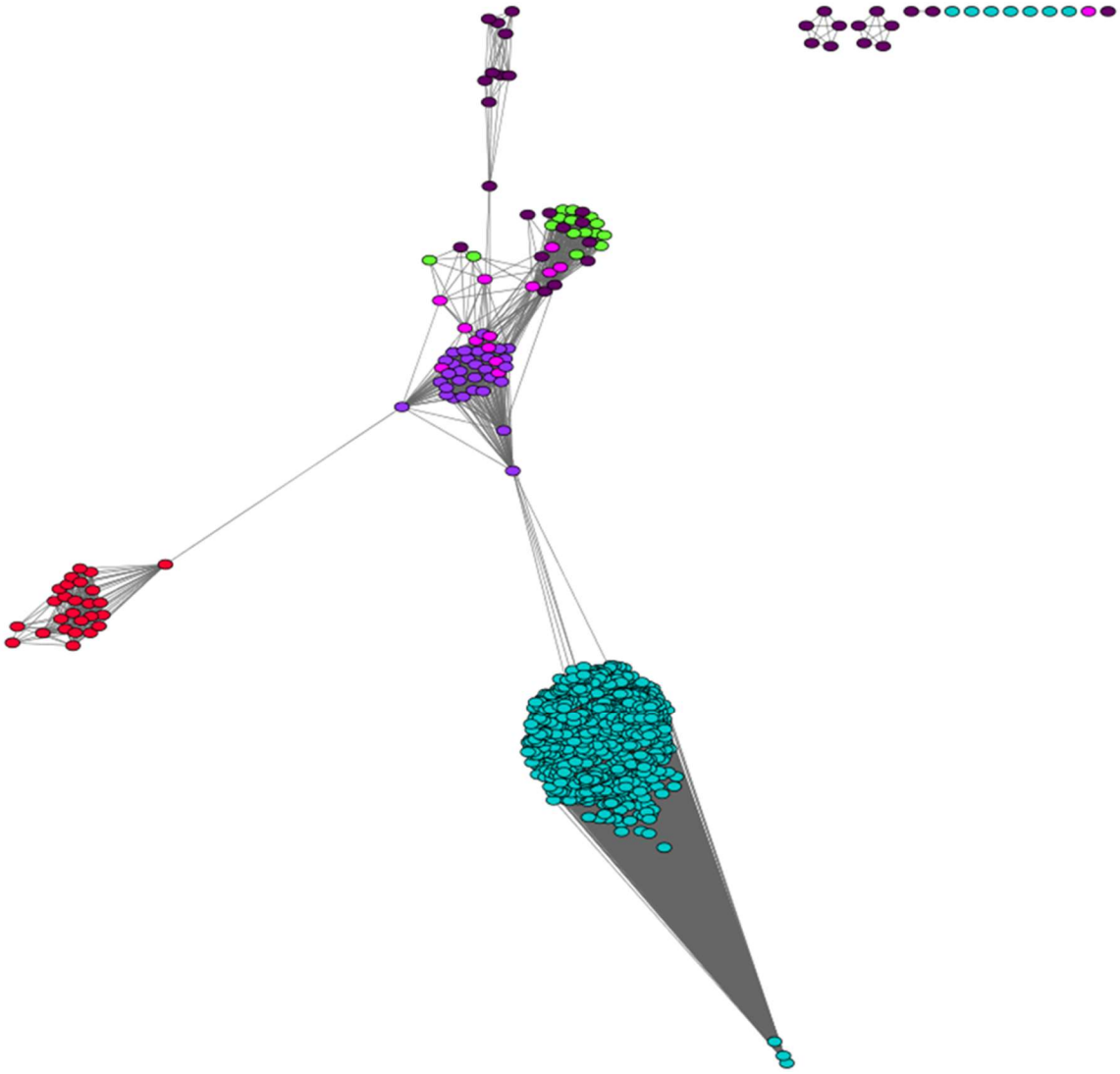

Figure 1. 8: Sequence similarity network. The network contains 838 nodes and 242,549 edges. Each node represents amino acid sequences that are 95% or more identical, and each edge connects a pair of sequences at an E-value of better than $1x10^{-30}$. This E-value is the one generated by the EFI-EST, and is not identical to E-values generated by BLAST. Teal: plants, indigo: actinobacteria, fuchsia: proteobacteria, green: cyanobacteria, red: fungi, plum: other.

account for the low observed yields. Specifically, low compatibility between MmFAMT and the fatty acid thioesterases used to generate the fatty acid substrate would result in methylation of only a fraction of the available fatty acids. The utilization of complementary combinations of thioesterases with mycobacterial methyltransferases that accept different fatty acid substrate

lengths can result in better yields. Based on our results, additional metabolic engineering, for example overexpression of an efficient SAH nucleosidase to avert product inhibition by SAH, can further increase the methyl ester production. These data provide a starting point for engineering efforts directed at exploiting both MmFAMT and homologs for the *in vivo* production of various fatty acid methyl esters of desired length.

1.5 References

1. *Energy Independence and Security Act of 2007 (EISA)*, in *Public Law 110-140*. 2007.

2. Naik, S.N., et al., *Production of first and second generation biofuels: A comprehensive review.* Renewable and Sustainable Energy Reviews, 2010. **14**(2): p. 578-597.

3. Demirbaş, A., *Diesel Fuel from Vegetable Oil via Transesterification and Soap Pyrolysis.* Energy Sources, 2002. **24**(9): p. 835-841.

4. Atadashi, I.M., M.K. Aroua, and A.A. Aziz, *High quality biodiesel and its diesel engine application: A review.* Renewable and Sustainable Energy Reviews, 2010. **14**(7): p. 1999-2008.

5. Leung, D.Y.C., X. Wu, and M.K.H. Leung, *A review on biodiesel production using catalyzed transesterification.* Applied Energy, 2010. **87**(4): p. 1083-1095.

6. Kung, Y., W. Runguphan, and J.D. Keasling, *From fields to fuels: recent advances in the microbial production of biofuels.* ACS Synthetic Biology, 2012. **1**(11): p. 498-513.

7. Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.* Nature, 2010. **463**(7280): p. 559-62.

8. Li, Q., W. Du, and D. Liu, *Perspectives of microbial oils for biodiesel production.* Applied Microbiology and Biotechnology, 2008. **80**(5): p. 749-56.

9. Nawabi, P., et al., *Engineering Escherichia coli for biodiesel production utilizing a bacterial fatty acid methyltransferase.* Applied and Environmental Microbiology, 2011. **77**(22): p. 8052-61.

10. Struck, A.W., et al., *S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications.* Chembiochem : a European Journal of Chemical Biology, 2012. **13**(18): p. 2642-55.

11. Blumenthal, R.M., X. Cheng, and E.B. Fauman, *Structure and evolution of AdoMet-dependent methyltransferases*. 1999, World Scientific Publishing Company Inc. p. 1-38.

12. Barry III, C.E., et al., *Mycolic acids: structure, biosynthesis and physiological functions.* Progress in Lipid Research, 1998. **37**(2–3): p. 143-179.

13. Huang, C.C., et al., *Crystal structures of mycolic acid cyclopropane synthases from Mycobacterium tuberculosis.* The Journal of Biological Chemistry, 2002. **277**(13): p. 11559-11569.

14. Yuan, Y., et al., *The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in Mycobacterium tuberculosis.* Molecular Microbiology, 1998. **29**(6): p. 1449-1458.

15. Yuan, Y. and C.E. Barry, *A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(23): p. 12828-12833.

16. Yuan, Y., et al., *Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(14): p. 6630-6634.

17. Glickman, M.S., *The mmaA2 gene of Mycobacterium tuberculosis encodes the distal cyclopropane synthase of the alpha-mycolic acid.* The Journal of Biological Chemistry, 2003. **278**(10): p. 7844-7849.

18. Glickman, M.S., J.S. Cox, and W.R. Jacobs Jr, *A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of Mycobacterium tuberculosis.* Molecular Cell, 2000. **5**(4): p. 717-727.

19. Boissier, F., et al., *Further insight into S-adenosylmethionine-dependent methyltransferases: structural characterization of Hma, an enzyme essential for the biosynthesis of oxygenated mycolic acids in Mycobacterium tuberculosis.* The Journal of Biological Chemistry, 2006. **281**(7): p. 4434-4445.

20. Akamatsu, Y. and J.H. Law, *The enzymatic synthesis of fatty acid methyl esters by carboxyl group alkylation.* Journal of Biological Chemistry, 1970. **245**(4): p. 709-13.

21. Orpiszewski, J., et al., *Multiple forms of O-methyltransferase involved in the microbial conversion of abietic acid into methyl abietate by Mycobacterium sp.* FEMS Microbiology Letters, 1991. **82**(2): p. 233-236.

22. Safayhi, H., M.I. Anazodo, and H.P. Ammon, *Calmodulin- and Ca2(+)-insensitive fatty acid methyltransferase from RINm5F cells. Inhibition by trifluoperazine and W7.* The International Journal of Biochemistry, 1991. **23**(7-8): p. 769-72.

23. Sastry, B.V., et al., *S-adenosyl-L-methionine-mediated enzymatic methylations in the rat retinal membranes.* Journal of Ocular Pharmacology, 1994. **10**(1): p. 253-63.

24. Liscombe, D.K., G.V. Louie, and J.P. Noel, *Architectures, mechanisms and molecular evolution of natural product methyltransferases.* Natural Product Reports, 2012. **29**(10): p. 1238-50.

25. McCarthy, A.A. and J.G. McCarthy, *The structure of two N-methyltransferases from the caffeine biosynthetic pathway.* Plant Physiology, 2007. **144**(2): p. 879-89.

26. Zubieta, C., et al., *Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family.* The Plant Cell, 2003. **15**(8): p. 1704-16.

27. D'Auria, J.C., F. Chen, and E. Pichersky, *Chapter eleven The SABATH family of MTS in Arabidopsis Thaliana and other plant species*, in *Recent Advances in Phytochemistry*, T.R. John, Editor. 2003, Elsevier. p. 253-283.

28. Doublie, S., *Preparation of selenomethionyl proteins for phase determination.* Methods in Enzymology, 1997. **276**: p. 523-30.

29. Sheldrick, G., *Experimental phasing with SHELXC/D/E: combining chain tracing with density modification.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(4): p. 479-485.

30. Bricogne, G., et al., *Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0.* Acta Crystallographica Section D, Biological Crystallography, 2003. **59**(Pt 11): p. 2023-30.

31. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D, Biological Crystallography, 2006. **62**(Pt 9): p. 1002-11.

32. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallographica Section D, Biological Crystallography, 2008. **64**(Pt 1): p. 83-9.

33. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 235-42.

34. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 355-67.

35. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(2): p. 213-221.

36. Langer, G., et al., *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.* Nature Protocols, 2008. **3**(7): p. 1171-9.

37. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(4): p. 486-501.

38. Lamzin, V.S. and K.S. Wilson, *Automated refinement of protein models.* Acta Crystallographica Section D, Biological Crystallography, 1993. **49**(Pt 1): p. 129-47.

39. Read, R.J., et al., *A new generation of crystallographic validation tools for the protein data bank.* Structure, 2011. **19**(10): p. 1395-412.

40. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(Pt 1): p. 12-21.

41. Dorgan, K.M., et al., *An enzyme-coupled continuous spectrophotometric assay for S-adenosylmethionine-dependent methyltransferases.* Analytical Biochemistry, 2006. **350**(2): p. 249-55.

42. Wooderchak, W.L., Z.S. Zhou, and J. Hevel, *Assays for S-adenosylmethionine (AdoMet/SAM)-dependent methyltransferases.* Current Protocols in Toxicology, 2008. **Chapter 4**: p. Unit4.26.

43. Hoye, T.R., C.S. Jeffrey, and F. Shao, *Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons.* Nature Protocols, 2007. **2**(10): p. 2451-8.

44. Gerlt, J.A., et al., *Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks.* Biochimica et Biophysica Acta-Proteins and Proteomics, 2015. **1854**(8): p. 1019-1037.

45. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Evolution, 1990. **215**(3): p. 403-10.

46. Altschul, S.F., *A protein alignment scoring system sensitive at all evolutionary distances.* Journal of Molecular Evolution, 1993. **36**(3): p. 290-300.

47. *UniProt: a hub for protein information.* Nucleic Acids Research, 2015. **43**(Database issue): p. D204-12.

48. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Research, 2003. **13**(11): p. 2498-504.

49. Choi-Rhee, E. and J.E. Cronan, *A nucleosidase required for in vivo function of the S-adenosyl-L-methionine radical enzyme, biotin synthase.* Chemistry and Biology, 2005. **12**(5): p. 589-93.

50. Kung, P.-P., et al., *SAH derived potent and selective EZH2 inhibitors.* Bioorganic and Medicinal Chemistry Letters, 2015. **25**(7): p. 1532-1537.

51. Obianyo, O. and P.R. Thompson, *Kinetic mechanism of protein arginine methyltransferase 6 (PRMT6).* Journal of Biological Chemistry, 2012. **287**(8): p. 6062-71.

52. Lee, W.J., J.Y. Shim, and B.T. Zhu, *Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids.* Molecular Pharmacology, 2005. **68**(4): p. 1018-30.

53. Zhao, N., et al., *Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family.* Plant Physiology, 2008. **146**(2): p. 455-67.

54. Bernstein, F.C., et al., *The Protein Data Bank: a computer-based archival file for macromolecular structures.* Journal of Molecular Biology, 1977. **112**(3): p. 535-42.

55. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Research, 2010. **38**(Web Server issue): p. W545-9.

56. Zubieta, C., et al., *Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases.* Nature Structural Biology, 2001. **8**(3): p. 271-9.

57. Zubieta, C., et al., *Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase.* The Plant Cell, 2002. **14**(6): p. 1265-77.

58. Alugupalli, S., F. Portaels, and L. Larsson, *Systematic study of the 3-hydroxy fatty acid composition of mycobacteria.* Journal of Bacteriology, 1994. **176**(10): p. 2962-2969.

59. Alugupalli, S., et al., *Chemical characterization of the ester-linked 3-hydroxy fatty acyl-containing lipids in Mycobacterium tuberculosis.* Journal of Bacteriology, 1995. **177**(15): p. 4566-4570.

60. Shimakata, T., Y. Fujita, and T. Kusaka, *Purification and Characterization of 3-Hydroxyacyl-CoA Dehydrogenase of Mycobacterium smegmatis.* The Journal of Biochemistry, 1979. **86**(5): p. 1191-1198.

61. Williams, K.J., et al., *The Mycobacterium tuberculosis β-oxidation genes echA5 and fadB3 are dispensable for growth in vitro and in vivo.* Tuberculosis (Edinburgh, Scotland), 2011. **91**(6-3): p. 549-555.

62. Sacco, E., et al., *The missing piece of the type II fatty acid synthase system from Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(37): p. 14628-14633.

63. Marrakchi, H., M.-A. Lanéelle, and M. Daffé, *Mycolic Acids: Structures, Biosynthesis, and Beyond.* Chemistry and Biology, 2014. **21**(1): p. 67-85.

CHAPTER 2: STRUCTURAL STUDIES OF THE WAX ESTER SYNTHASE FROM *Marinobacter aquaeolei* VT8

2.1 Introduction

Wax esters have widely been used in the cosmetics, lubricant, candle, and food industries since the 19th century [1]. Up until the 1960s, the main source of wax esters was oil from sperm whales, until their endangerment resulted in a global ban on whaling [2]. Since the Endangered Species Act in 1973 and the ban on sperm oil use, alternative sources of wax esters have been sought. Currently, wax esters are extracted from plants (e.g. jojoba oil) or are chemically synthesized from expensive precursors [3, 4]. Hence, there is a strong demand for the development of large-scale processes for the production of low cost wax esters. Towards that end, there has been increasing focus in the development of methods for the enzymatic biosynthesis of wax esters. Current methods center on the use of enzymes immobilized on solid supports, as well as the use of cellular 'factories' engineered to produce the desired wax esters [4-8].

Wax esters are esters of long chain fatty acids and long chain alcohols that are formed *in vivo* by the condensation of a fatty alcohol and a fatty acyl-Coenzyme A (acyl-CoA), a reaction catalyzed by acyltransferases termed wax ester synthases (WS, EC 2.3.1.75). These acyltransferases are largely found in bacteria, plants and animals, and can also catalyze the transesterification of acyl-CoAs with diacylglycerols (DAGs), an activity known as diacylglycerol *O*-acyltransferase (DGAT, EC 2.3.1.20) [9-13]. In bacteria, all the WSs that have been characterized thus far are bifunctional, and encode both WS and DGAT activities [13, 14]. Although a few bifunctional WS/DGATs from plants and marine protists have been characterized,

there is dearth in information about the large number of predicted WS/DGAT sequences in eukaryotic genomes that have yet to be experimentally validated [14-18].

Bacterial WS/DGATs are presumptively promiscuous enzymes as homologs from different species show preference for different acyl chain length and/or degree of saturation [19-21]. This promiscuity has been exploited in the biotechnological utilization of different WS/DGAT homologs for the *in vivo* production of wax esters and biodiesel (esters consisting of medium chain fatty acids and ethanol) [5, 8, 22, 23]. Currently, the most well studied WS/DGATs are those from *Acinetobacter baylyi* ADP1 (Ab-WS/DGAT) and *Marinobacter aquaeolei* VT8 (Ma-WS/DGAT) [12, 14, 19, 20, 24, 25]. These enzymes have been utilized for the *in vivo* production of biodiesel and other wax esters by several laboratories, and engineering efforts to alter their substrate specificity have been an area of significant focus [21, 24, 26-32]. There is currently no structure available for any member of this enzyme family in the Protein Data Bank (PDB), which limits engineering efforts aimed at altering substrate scope for a given WS/DGAT.

Bacterial WS/DGATs share low overall sequence similarity not only with eukaryotic WS/DGATs, but also with other bacterial homologs, both across and within species [13, 14]. Despite the overall low sequence similarity, all WS/DGATs share a conserved HHXXXDG motif, which is also found in other acyltransferases, such as the condensation domains of non-ribosomal peptide synthases, chloramphenicol acetyltransferase, and polyketide-associated proteins (Pap) from Mycobacteria [33, 34]. In each of these acyltransferases, the conserved motif contains a catalytic His that has been proposed to act as the general base that abstracts a proton from the alcoholic substrate hydroxyl to generate a nucleophilic alkoxide (Figure 2. 1) [33-36]. Attack of the alkoxide onto the carbonyl carbon of the acyl-CoA thioester produces a tetrahedral intermediate, which subsequently collapses to yield the acylated alcohol and free CoA. In the case

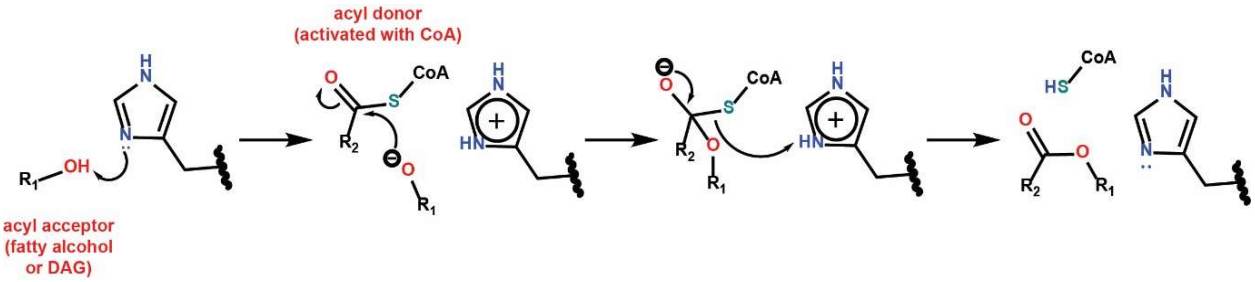

Figure 2. 1: Proposed mechanism of bacterial WS/DGAT. A catalytic histidine of the conserved motif HHXXXDG abstracts a proton from the hydroxyl moiety of the substrate generating an alkoxide ion. This alkoxide ion subsequently attacks the carbonyl carbon of the thioester bond in acyl-CoA resulting in a tetrahedral intermediate which collapses resulting in the acylated product and free CoA. Abbreviations: DAG for diacylglycerol and CoA for Coenzyme A.

of chloramphenicol acetyltransferase and PapA5 from *Mycobacterium tuberculosis*, the second His in the HHXXXDG motif is proposed to be the general base [34-36]. However, in bacterial WS/DGATs, the identity of the catalytic His has yet to be established. *In vitro* studies and mutational analyses performed with WS/DGATs from *Acinetobacter baylyi* ADP1 and *Marinobacter aquaeolei* VT8 showed that mutation of either of the two His severely affected the activity of the enzyme, suggesting that both are important for catalysis [13, 14]. Mutation of the conserved Asp in this motif, decreased, but did not abolish the activity of the enzyme suggesting a structural, rather than catalytic role [14, 37].

Here, we report the first crystal structure of any WS/DGAT, that from *Marinobacter aquaeolei* VT8 (Ma-WS/DGAT, homolog WS1, UniProt accession number A1TX06). The structural data in conjunction with structure-guided mutational analysis allowed assignment of the roles for residues in the conserved HHXXXDG motif, and provide a molecular framework for identifying the binding sites for each of the two substrates. We also utilized the structure to guide engineering experiments aimed at altering the substrate scope of the enzyme. Our studies demonstrate that structure-guided engineering may be an expedient route for producing

WS/DGAT variants with different product profiles, which is an area of immense interest in biotechnology.

2.2 Materials and Methods

Chemical and Reagents

All chemical reagents were purchased from Sigma-Aldrich unless otherwise noted. All of the material used for protein production and purification were purchased from GE Healthcare.

Cloning and Site-Specific Mutagenesis

Ma-WS/DGAT (GenBank accession number WP_011783747 – locus NC_008740.1) was amplified using the PCR with template genomic DNA of *Marinobacter aquaeolei* VT8 (DSM 11845) and primers designed based on the published sequence. The gene was inserted into the pET His6-Sumo-TEV LIC cloning vector (2S-T) (Addgene plasmid # 29711), and this plasmid was used as template for the generation of the site-specific mutants by PCR (primers shown in Table 2. 1). Sanger sequencing was used to confirm the integrity of all recombinant plasmids (ACGT, Inc.).

Protein Expression and Purification

Expression vectors bearing wild type or mutant Ma-WS/DGAT were transformed into *E. coli* Rosetta 2(DE3) cells for heterologous protein production. A 4 ml fresh starter culture was inoculated in 1 L Luria-Bertani growth medium supplemented with 100 µg/mL ampicillin and 25 µg/mL chloramphenicol. The culture was grown at 37 °C until the absorbance at 600 nm reached 0.6 – 0.8, at which point protein production was induced by addition of 0.3 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG). The culture was then cooled to 18 °C and grown with shaking for an additional 18 h. Cells were collected by centrifugation, resuspended in Buffer A (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol), and lysed by multiple passages through a C5 Emulsiflex (Avestin) cell homogenizer. Following centrifugation of the lysate, the supernatant was applied to a 5 mL His-Trap (GE Biosciences) column that was previously equilibrated with Buffer B (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 30 mM imidazole, 10% glycerol). The column was washed with 20 mL of buffer B, 20 mL of buffer B supplemented with 5 mM $MgCl_2$ and 5 mM ATP, and 15 mL of buffer C (20 mM Tris-HCl, pH 8.0, 1M NaCl, 50 mM imidazole, 10% glycerol). Elution of the bound, His6-SUMO-tagged protein was carried out using a gradient of increasing imidazole concentration (to a final concentration of 200 mM). Fractions containing protein of the highest purity (as determined by sodium dodecyl sulphate - polyacrylamide electrophoresis, SDS-PAGE)

Table 2. 1: Primers used for amplification of wild type Ma-WS/DGAT and generation of mutants.

Primer name	Sequence (5'-3')
MnWS/DGAT_F	TACTTCCAATCCAATGCA ATGACGCCCCTGAATCCCACTG
MnWS/DGAT_R	TTATCCACTTCCAATGTTATTATTACAGACCGGCGTTGAGCTCC
MnWS/DGAT_D8A_F	TACTTCCAATCCAATGCA ATGACGCCCCTGAATCCCACTGCCCAGCTC
MnWS/DGAT_H136N_F	CTCTACACCAAGGTTCACAATTCCCTGGTGGACGGT
MnWS/DGAT_H136N_R	ACCGTCCACCAGGGAATTGTGAACCTTGGTGTAGAG
MnWS/DGAT_D140A_F	CACCATTCCCTGGTGGCCGGTGTCTCGGCCATG
MnWS/DGAT_D140A_R	CATGGCCGAGACACCGGCCACCAGGGAATGGTG
MnWS/DGAT_G25V_F	CCCATGCATGTGGGCGTCCTCCAGCTGTTTTCC
MnWS/DGAT_G25V_R	GGAAAACAGCTGGAGGACGCCCACATGCATGGG
MnWS/DGAT_A144V_F	GTGGACGGTGTCTCGGTCATGCGCATGGCCACC
MnWS/DGAT_A144V_R	GGTGGCCATGCGCATGACCGAGACACCGTCCAC
MnWS/DGAT_A144F_F	GTGGACGGTGTCTCGTTCATGCGCATGGCCACC
MnWS/DGAT_A144F_R	GGTGGCCATGCGCATGAACGAGACACCGTCCAC

were pooled, and the His_6-SUMO-tag was removed by overnight incubation with tobacco etch viral protease, followed by dialysis against Buffer D (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1mM DTT, 10% glycerol) at 4 °C. The dialyzed samples were re-applied to a 5 ml His-Trap (GE Biosciences) column to remove the His6-SUMO-tag, and pure fractions of the tag-excised protein were further purified using size exclusion chromatography (SEC) (Superdex Hiload 200 16/60) in a buffer of 20 mM Tris-HCl, pH 8.0, 1 M NaCl. Samples were concentrated using a 10,000 Da molecular weight cutoff Amicon centrifugal filters. Selenomethionine-labeled (SeMet) Ma-WS/DGAT was produced by repression of methionine biosynthesis in defined media prior to protein induction, and was purified as described above. All proteins were flash frozen in liquid nitrogen and stored at -80 °C.

For activity studies, Ma-WS/DGAT wild type and mutants were purified as described above with the following modification: After cleavage of the His_6-SUMO-tag and the second His-affinity column, all samples were passed through a PD-10 desalting column (GE Healthcare), instead of SEC, for buffer exchange (20 mM Tris-HCl, pH 8.0, 1 M NaCl). The purity of all samples was assessed by SDS - PAGE analysis (Figure S1), and protein concentrations were determined by NanoDrop (Thermo Fisher Scientific) analyses using the molecular weights and extinction coefficients calculated by the ProtParam tool (ExPASY server) (Table 2. 2) [38].

Table 2. 2: Extinction coefficients and molecular weights as calculated by ProtParam (ExPASY server) [38] and used for determination of the concentrations of wild type Ma-WS/DGAT and mutants.

Ma-WS/DGAT variant	ε_{280} ($M^{-1}cm^{-1}$) x 10^3	MW (kDa)
WT	61.88	51.75
D8A	61.88	51.70
G25V	61.88	51.79
H136N	61.88	51.72
D140A	61.88	51.70
A144V	61.88	51.77
A144F	61.88	51.83

Crystallization

Initial crystallization conditions were determined by the sparse matrix sampling method using commercial screens. Crystals of Ma-WS/DGAT were grown using the hanging vapor drop diffusion method. Briefly, 1 µl of protein at 4 mg/ml concentration was incubated with 1.2 molar equivalents of palmitoyl-CoA (C16-CoA), mixed with 2 µl of precipitant solution (0.2 M L-Pro, 0.1 M MES, pH 7.0, 8% PEG 3350), and equilibrated over a well containing the same precipitant solution at 9 °C. Crystals took 1 month to grow, and were used to seed subsequent optimization trays. After several rounds of optimization and seeding, diffraction-quality crystals were obtained at 6 mg/ml protein concentration in 0.2 M L-Pro, 0.1 M MES, pH 7.0, 12% PEG 3350 at 16 °C. The crystals were sequentially soaked in precipitant solution supplemented with increasing concentrations of 10, 20 and 30% ethylene glycol prior to vitrification by direct immersion in liquid nitrogen. SeMet crystals were grown by seeding with seeds from native crystals at 2 mg/ml protein concentration in 0.2 M L-Pro, 0.1 M MES, pH 7.0, 12% PEG 3350 at 16 °C. Attempts to

soak the crystals with palmitoyl-CoA (C16-CoA), hexanoyl-CoA (C6-CoA), CoA and hexanol (C6-OH) were not successful.

Data Collection, Phasing and Structure Determination

X-ray diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT), Sector 21, Argonne National Laboratory. All data were indexed, integrated and scaled using either HKL2000 or AutoProc [39-42]. Initial crystallographic phases were determined using single wavelength anomalous diffraction data collected from crystals of SeMet-labeled Ma-WS/DGAT at the Se absorption edge (λ=0.97872 Å) to a diffraction limit of 2.9 Å. Heavy atom sites were located using the Hybrid Substructure Search submodule of the Phenix program suite [43-45], and were subsequently imported into the AutoSol wizard of Phenix for maximum likelihood phasing [46], yielding an initial figure of merit of 0.369. A partial model of the structure was generated using automated methods as implemented in RESOLVE, and this initial model was further extended using the Phenix Autobuild wizard [47]. The resultant model was used as search probe for the determination of the native crystal structure by molecular replacement as implemented in the Phenix program suite [45]. The resultant solutions were subsequently used as starting models for several rounds of automated model building using Phenix Autobuild and Buccaneer [47-50], followed by rounds of manual rebuilding using Coot [51], and refinement using either Phenix Refine or REFMAC5 [52, 53]. Water molecules were added during refinement with Phenix Refine, and confirmed by manual inspection. In all cases, the quality of the in-progress model was

routinely monitored using both the free R factor and MolProbity for quality assurance [45, 54]. All crystallographic data collection and refinement statistics are provided in Table 2. 3.

Table 2. 3: Data collection, phasing and refinement statistics (1. Highest resolution shell is shown in parenthesis, 2. $R_{sym} = \Sigma |(I_i - \langle I_i \rangle| \Sigma I_i$ where I_i = intensity of the ith reflection and $\langle I_i \rangle$ = mean intensity, 3. Mean figure of merit, 4. R-factor = $\Sigma (|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement).

	SeMet Ma-WS/DGAT	**Native Ma-WS/DGAT**
Data collection		
Space Group	P21	P21
a, b, c (Å), β (°)	69.4, 103.7, 69.9, 104.1	69.3, 103.9, 69.5, 103.5
Resolution (Å)	103.7-2.9	103.9-2.4
R_{sym} (%)[1, 2]	14.9 (82.7)	6.4 (73)
I/σ(I)	13.1 (3.2)	23.4 (3.3)
Completeness (%)	100 (100)	100 (99.9)
Redundancy	9.7 (9.9)	10.8 (10.9)
Total reflections	195,392	391,121
Unique reflections	20,209	36,232
Phasing		
Figure of Merit[3]	0.369	
Refinement		
Resolution (Å)		51.9-2.4
No. reflections used		36,196
R_{work} / R_{free}[4]		0.21/0.28
Number of atoms		
Protein		6425
Water		42
B-factors		
Protein		64.60
Water		55.88
R.m.s deviations		
Bond lengths (Å)		0.008
Bond angles (°)		0.992
MOLPROBITY statistics		
Clash score		6.30
Ramachandran outliers/ allowed/ favored (%)		0.89/4.56/94.55

Wax Ester Synthase Activity Assay

The wax ester synthase (WES) activity of the wild type enzyme and mutants was investigated by gas chromatography-mass spectroscopy (GC/MS). 200 µl reactions containing 5.0 µM Ma-WS/DGAT, 50 µM palmitoyl-CoA (C16-CoA) or 100 µM hexanoyl-CoA (C6-CoA) and 1 mM 1-hexanol (C6-OH) were incubated at 30 °C for 0.5 h and stopped by addition of 200 µL chloroform. They were subsequently vortexed and span down, and the chloroform phase was extracted and analyzed by GC/MS by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign). All reactions were performed in 50 mM NaP_i, pH 7.5, 300 mM NaCl and 1.45% (final concentration) dimethyl sulfoxide (DMSO). The activity of the mutants was expressed as percentage of WE formation in respect to the wild type reaction (considered as 100%). For this purpose, the area under the WE peak of the wild type reaction was considered as 100%, and the areas under the WE peak of the mutants were converted to % of activity with the following equation: % of activity = (peak area of mutant) x 100 / (peak area of wild type). Control reactions without addition of the enzyme were also analyzed. Results are means of duplicate experiments.

For the competition assay, reactions were prepared as described above with the following modification: 50 µM C16-CoA and 50 µM C6-CoA were both added in the reactions and the reactions were initiated by addition of 1 mM hexanol. The sum of the areas under the peaks for hexyl-palmitate and hexyl-caproate was considered as the total amount of wax esters produced (100%) and the relative abundance of each peak is reported as % of hexyl-palmitate and % of hexyl-caproate produced by each enzyme. Results are means of duplicate experiments.

Diacylglycerol Acyltransferase Activity

The diacylglycerol acyltransferase (DGAT) activity of the wild type enzyme was investigated by liquid chromatography-mass spectroscopy (LC/MS). 200 µl reactions containing 5.0 µM Ma-WS/DGAT, 50 µM palmitoyl-CoA (C16-CoA) and 200 µM 1,2-dipalmitoyl-sn-glycerol were incubated at 30 °C for 1 h and stopped by addition of 200 µL chloroform. They were subsequently vortexed and span down, and 150 µL of the chloroform phase were extracted, evaporated, resuspended in 20 µL chloroform and analyzed by LC/MS by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign). All reactions were performed in 50 mM NaP_i, pH 7.0, 300 mM NaCl and 1.45% (final concentration) dimethyl sulfoxide (DMSO).

Gas Chromatography-Mass Spectrometry (GC/MS)

The samples were analyzed in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, using a GC/MS system (Agilent Inc, CA, USA) consisting of an Agilent 7890 gas chromatograph, an Agilent 5975 MSD and an Agilent 7683B autosampler. Gas chromatography was performed on a ZB-1MS (30mm×0.32mm I.D. and 0.25µm film thickness) capillary column (Phenomenex, CA, USA). The inlet and MS interface temperatures were set to 2800 °C., and the ion source temperature was adjusted to 230 °C. An aliquot of 1 µL was injected in a splitless mode (purge time 0.5 min). The helium carrier gas was kept at a constant flow rate of 2.3 mL/min. The temperature program was: 2 min isothermal heating at 180 °C, followed by an oven temperature increase of 15 °C min^{-1} to 330 °C and a final 10 min at 330 °C. The mass spectrometer was operated in positive electron impact mode (EI) at 69.9 eV ionization energy at m/z 33-500 range. Mass spectra were recorded in the combined scan/SIM

mode. For SIM mode m/z 340, 257 (hexyl-palmitate) and 117 (hexyl-caproate) were tracked. Hexyl-caproate was confirmed with NIST08 commercial database (NIST, MD, USA) while hexyl-palmitate was confirmed by high resolution GC/MS. The mass spectrum of target compound was evaluated using the MSD Chemstation E.02.01.1177 software (Agilent, Palo Alto, CA, USA).

Liquid Chromatography-Mass Spectrometry (LC/MS)

The sample was analyzed by Q-Exactive MS system (Thermo. Bremen, Germany) in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Software Xcalibur 4.1.31.9 was used for data acquisition and analysis. Mass spectrum was acquired under positive ESI with the following settings: sheath gas flow rate, 49; aux gas flow rate: 12; sweep gas flow rate, 2; spray voltage, 3.5 kV; capillary temp, 259 °C; Aux gas heater temp, 419 °C. The resolution was set to 70,000. The AGC target was 1E6 with a maximum injection time of 50 ms.

Amino Acid Residues Conservation Analysis

The conservation of amino acids in Ma-WS/DGAT was analyzed by utilizing the ConSurf server [55, 56], a bioinformatics tool that analyzes the conservation of amino acids in a protein taking into account the evolutionary relationships between homologs of the same family. The homologs were collected from UNIREF90 using PSI-BLAST as the search algorithm. The settings were as follows: 90% maximum identity between sequences, 35% minimum identity between sequences, total of 500 closest sequences. The retrieved sequences were aligned using ClustalW, and a Bayesian method was used for the calculation of the conservation scores. There were 419 PSI-BLAST hits, but only 396 of which were unique, and were used for the calculation.

Sequence Logo

A sequence logo was generated by WebLogo (http://weblogo.berkeley.edu/) [57, 58] using as input the curated seed alignment of PFAM for wax synthases (PFAM family PF03007) [59]. The numbering of the residues corresponds to the sequence of Ma-WS/DGAT.

2.3 Results and Discussion

Ma-WS/DGAT activity

The WS/DGAT from *Marinobacter aquaeolei* VT8 (Ma-WS/DGAT) studied here has already been studied in terms of alcohol substrate promiscuity [20, 21]. However, before moving forward to structural studies, we confirmed the WS and DGAT activities of the enzyme using recombinant purified Ma-WS/DGAT. The enzyme was able to produce hexyl-palmitate and tripalmitate when palmitoyl-CoA and hexanol or 1,2-sn-dipalmitate were provided respectively (Figure 2. 2).

Crystal Structure of Ma-WS/DGAT

Ma-WS/DGAT was crystallized as a dimer at 2.4 Å in space group $P2_1$ showing a structure consisting of an N-terminal and C-terminal domains, connected via a helical linker (Figure 2. 3A). The N-terminal domain is composed of a mixed β-sheet (strands β1, β4, β5, β6, β11) flanked by 4 α-helices (helices α1-α4), and a small antiparallel β-sheet (strands β2 and β3). The C-terminal domain consists of a mixed β-sheet (strands β7-β10 and β12-β13) that on one side is covered by 4 α-helices (α6, α7, α8, α10), and on the other side it faces helix α9 and the N-terminal domain. The two domains are connected by helix α5 (residues 200-214). Disordered regions with poor electron density were built with caution, however, there are still small regions of unassigned density. The

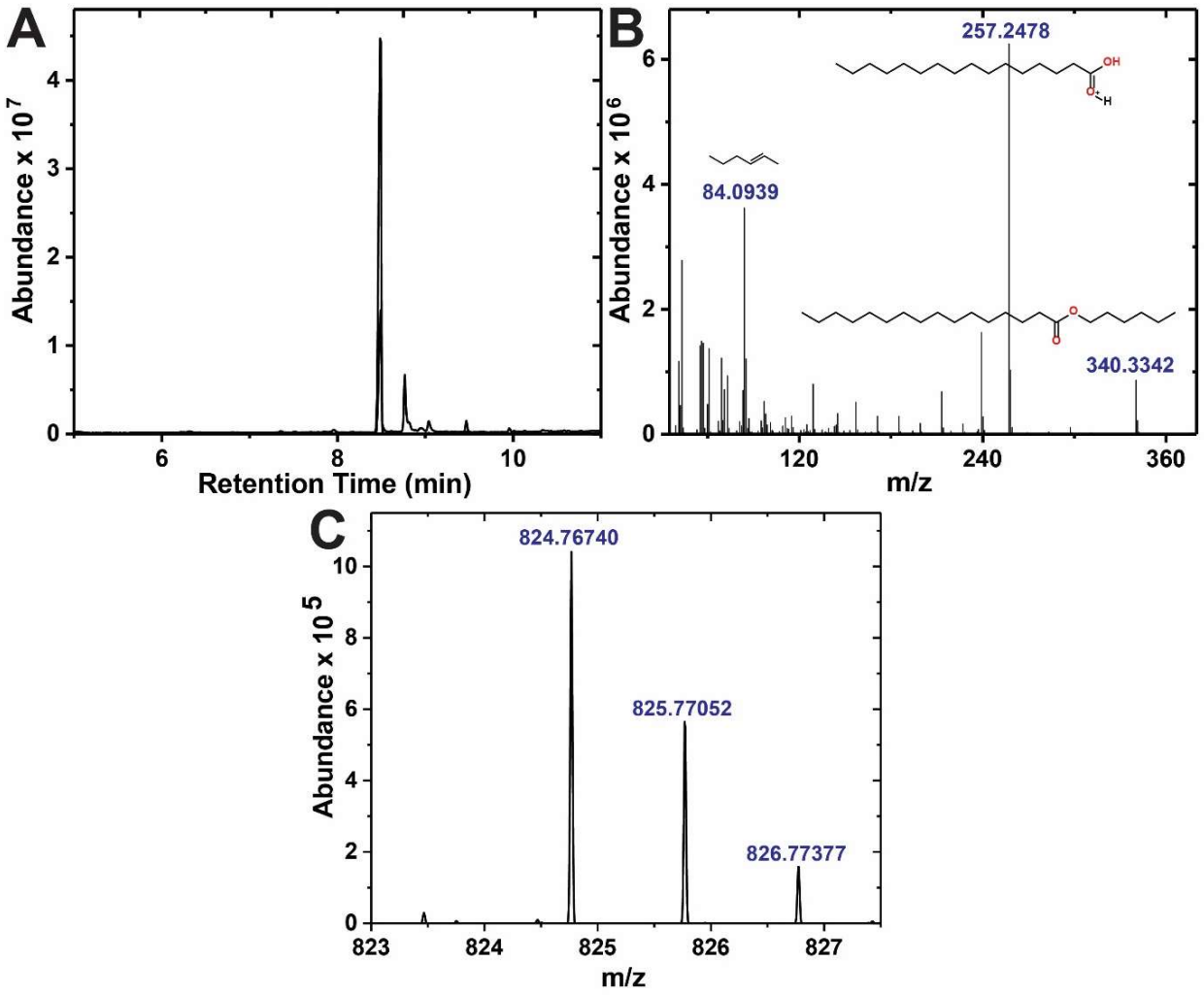

Figure 2. 2: (A) GC/MS chromatogram of hexyl-palmitate. (B) Fragmentation pattern of peak at 8.4 min. Only the high-resolution masses of the peaks corresponding to hexyl-palmitate and its fragments are shown. (C) High-resolution mass spectrum confirms target compound (tripalmitate) in the form of $[M + NH_4]^+$ (m/z = 824.76740). Such NH_4^+ adduct is very common for tripalmitate.

final model is missing residues 169-198, 219-225, 296-308 and 341-348 from chain A, and residues 169-190, 217-225 and 296-308 from chain B. This high degree of disorder and flexibility may provide one reason why the crystals took long to grow (about one month).

The crystal structure reveals the presence of a long cavity that is formed at the core of the N-terminal domain (pocket 1) and extends to the interface of the two domains (pocket 2) (Figure 2. 3B). The conserved HHXXXDG motif is located on a loop that links strand β6 with helix α4

with His136 positioned to face towards pocket 2, while His135 is directed away from this cavity (Figure 2. 3A-C). Based on the orientation of the two conserved His, it is likely that His136 is catalytic while His135 plays a structural role. This notion is further supported by the fact that His135 is within hydrogen bonding distance from the highly conserved Glu15, the side chain of Ser137 and the amide nitrogens of His136 and Ser137 (Figure 2. 3C). Disruption of these interactions would explain why, in prior studies, enzyme activity was severely affected when the first His was substituted by Leu in the WS/DGAT from *Acinetobacter baylyi* ADP1 or by Ala in a different WS/DGAT homolog from *Marinobacter aquaeolei* VT8 (WS2, UniProt accession number A1U572) [13, 14]. In addition to His135, the crystal structure suggests that Asp140 also play a structural role. The side chain of Asp140 is involved in hydrogen bonding interactions with the first three residues of helix α4 (Figure 2. 3D). Presumably, disruption of these interactions would affect the structural integrity of the loop where the catalytic His135 is located and would account for the significant loss of activity when by an Ala substitution at this conserved Asp [14, 37].

In one of the two chains in the non-crystallographic dimer, there is unassigned electron density adjacent to His136. Attempts to fit glycerol, MES and PEG were not successful as these ligands were too small to justify the observed density. It is possible that some other ligand co-purified with the enzyme, but the quality of the density is not sufficient for a conclusive assignment.

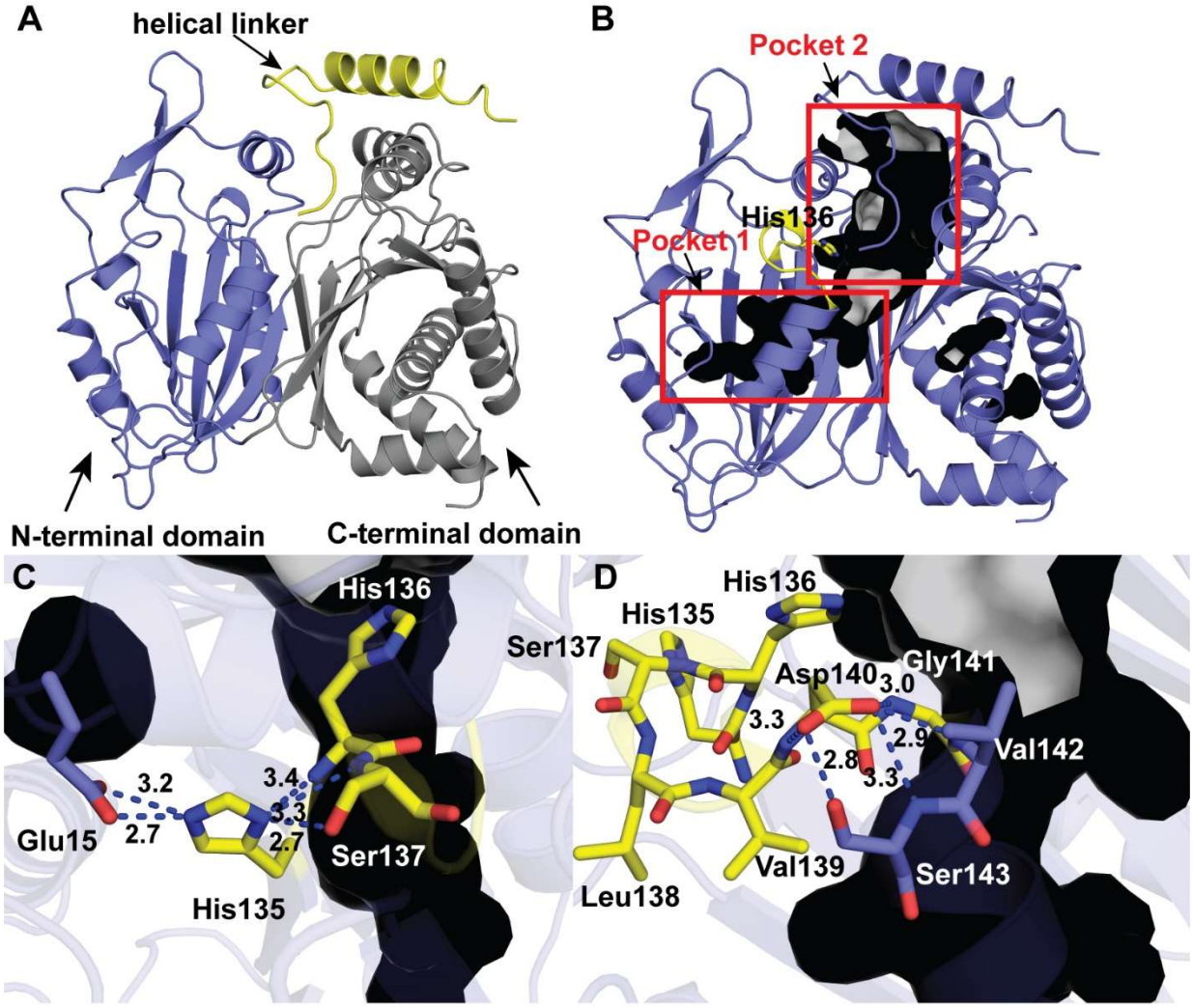

Figure 2. 3: (A) Ribbon diagram of the overall structure of Ma-WS/DGAT (one monomer shown). The N-terminal domain is shown in purple, the C-terminal domain in gray and the helical linker in yellow. (B) The cavity of Ma-WS/DGAT is shown in black, and the conserved motif HHXXXDG (HHSLVDG in Ma-WS/DGAT) is colored in yellow. For description purposes, the cavity is divided into pocket 1 and pocket 2. (C) His135 points away from the enzyme cavity, and is in hydrogen bonding distance from Glu15, His136 and Ser137. All 4 residues are highly conserved. (D) Asp140 seems to play structural role forming interactions mainly with the peptide backbone. Distances shown are in Å.

.

Amino Acid Residues Conservation

The evolutionary conservation of amino acids across WS/DGATs was calculated and mapped onto the crystal structure of Ma-WS/DGAT. This amino acid conservation can provide useful information about functionally important residues, especially in protein families like the WS/DGAT that include homologs which share a low overall sequence similarity. In total, 396 sequences with sequence identity of 35-90% to Ma-WS/DGAT were aligned, and the conservation score as calculated by ConSurf server [55, 56] was mapped on the crystal structure of Ma-WS/DGAT (Figure 2. 4). Most of the highly conserved residues are located on regions that are important for the structural integrity of the enzyme. In addition to the conserved motif, only a few highly conserved residues line the enzyme cavity. These residues are located in a "ring" (Figure 2) around where the catalysis is expected to take place and may be involved in orienting the substrates and stabilizing the tetrahedral intermediate. This low conservation of the cavity lining amino acid residues could justify the differences in substrate specificities between different

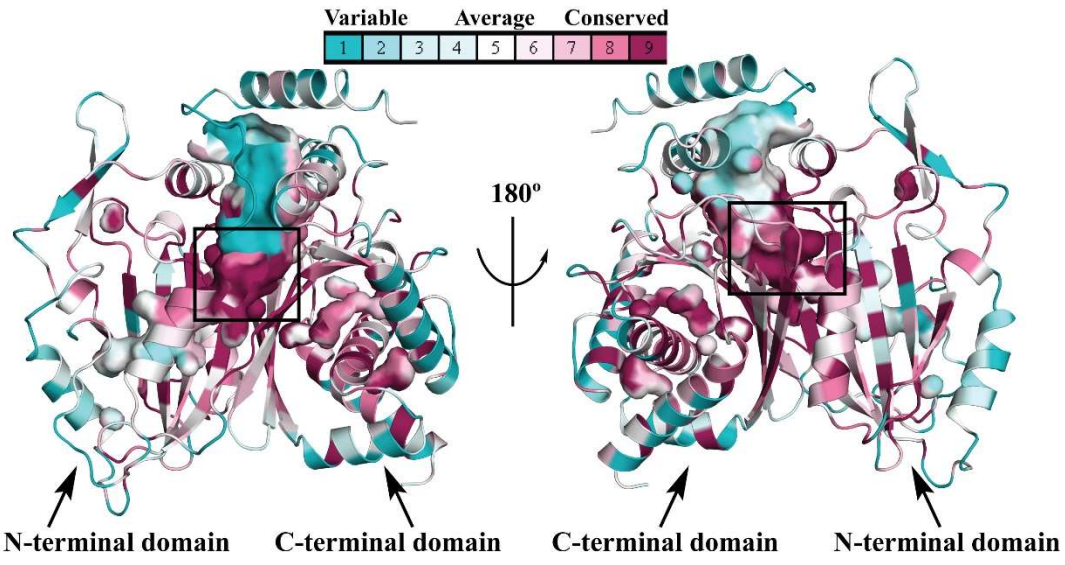

Figure 2. 4: Amino acid sequence conservation mapped onto the structure of Ma-WS/DGAT (as calculated by ConSurf). The rectangle indicates a "ring" of highly conserved residues around the active site.

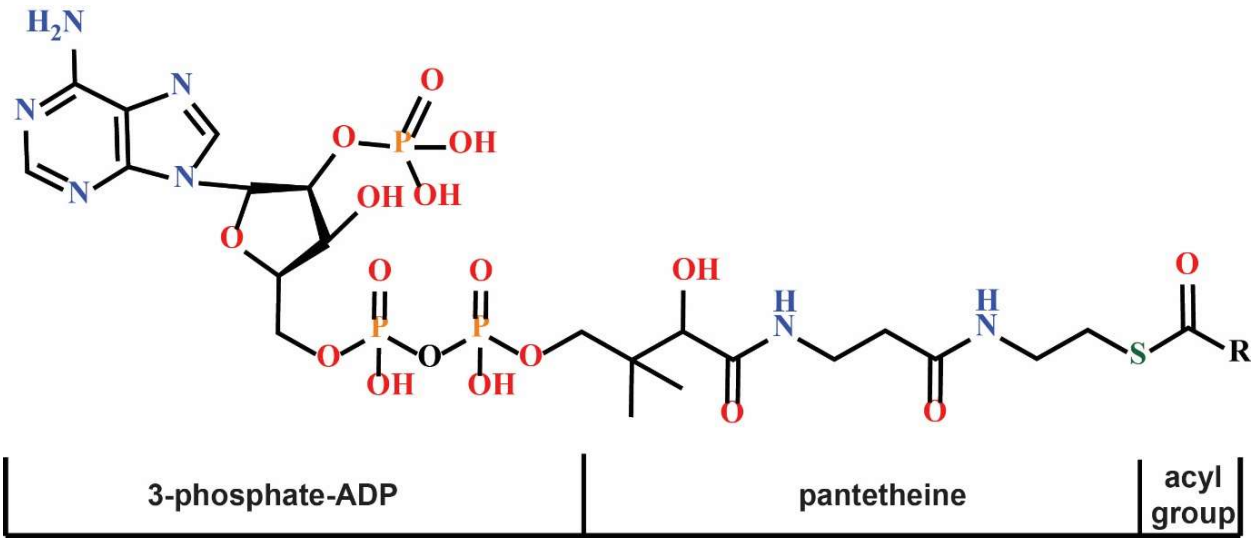

Figure 2. 5: Schematic representation of Coenzyme A and its building blocks (ADP for adenosine diphosphate).

homologs and could be exploited for structure-based engineering of Ma-WS/DGAT towards the design of variants with different substrate specificities.

Substrate Binding Pockets

The orientation of the conserved motif HHXXXDG suggests that the fatty alcohol or DAG substrates bind either in pocket 1 or pocket 2. Structural comparison of Ma-WS/DGAT with other acyltransferases, such as the TRI3 trichothecene 15-*O*-acetyltransferase from *Fusarium sporotrichioides* (FsTRI3) (PDB ID 3FP0, Z-score 18.9 and RMSD 4.6 Å as calculated by the Dali server) (Figure 2. 6A) [60, 61], suggests that the acyl acceptor substrates bind in pocket 2. This is in agreement with mutational studies where mutations at residues that reside in pocket 2 altered the alcohol substrate specificity of Ma-WS/DGAT [24]. In these studies, substitutions at Leu356 and Met405 with either larger or smaller hydrophobic residues resulted in Ma-WS/DGAT variants with different selectivity towards small, branched and aromatic alcohols. It is notable that both

residues are not highly conserved; in nature, there are WS/DGAT homologs that have several different amino acids in these positions (4-5 conservation score in the scale of ConSurf) [55].

Further structural comparison of Ma-WS/DGAT with other acyltransferases suggests that the phosphoadenosine moeity of acyl-CoA (Figure 2. 5) is exposed to the solvent and the pantetheine portion is extended into the active site with the thioester positioned close to His136 and the acyl chain buried in pocket 1 (Figure 2. 6C). In this binding mode, the acyl chain is shielded from the solvent and the thioester is positioned next to the catalytic His136 and the acyl acceptor.

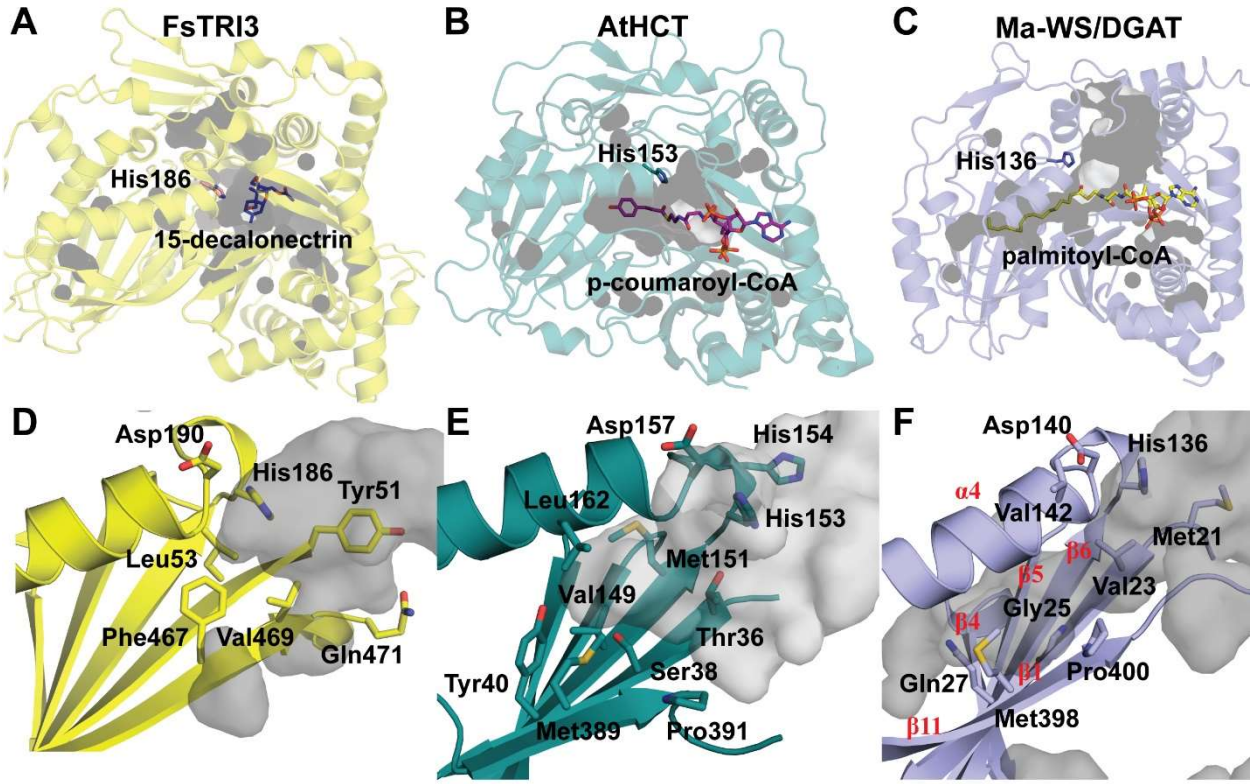

Figure 2. 6: Ribbon diagram and cavity of: (A) TRI3 trichothecene 15-O-acetyltransferase from *Fusarium sporotrichioides* (FsTRI3) (PDB ID 3FP0).[61] The substrate (acetyl acceptor) is shown in sticks. (B) Hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase from *Arabidopsis thaliana* (AtHCT) (PDB ID 5KJT).[62] The Coenzyme A (CoA) activated substrate (p-coumaroyl-CoA) is shown in sticks. (C) Ma-WS/DGAT with palmitoyl-CoA modeled based on the superposition with AtHCT. The catalytic histidine residues are shown in sticks. (D, E, F) Comparison of the residues that line and constrict the pocket at the core of the N-terminal domains of FsTRI3, AtHCT and Ma-WS/DGAT respectively.

Since pocket 1 is about 24 Å long, Ma-WS/DGAT could accommodate acyl chains of 19-20 carbons long if bound in a fully extended conformation. In comparison with other CoA dependent transferases that share the same fold, the pocket formed at the core of the N-terminal domain reflects the size and shape of the CoA activated substrate. In the active site of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase from *Arabidopsis thaliana* (AtHCT) (PDB ID 5KJT, Z-score 16.0 and RMSD 4.9 Å as calculated by the Dali server) [60, 62], *p*-coumaroyl-CoA is bound in a similar mode, and the size of the pocket is sufficient to accommodate the cognate ligand (Figure 2. 6B). Likewise, FsTRI3 acetyltransferase [61] lacks a pocket at the core of the N-terminal domain as the substrate acetyl group is small and fits in the pocket formed at the interface of the two domains (Figure 2. 6A).

Superposition of FsTRI3 and AtHCT with Ma-WS/DGAT reveals the molecular features that dictate the formation, size and shape of pocket 1. These features include shortening of strand β11 by introduction of a Pro residue at the C-terminal end of the strand, introduction of bulky hydrophobic residues such as Phe467 in FsTRI3 and Tyr40 in AtHCT to define the length of the binding pocket, and use of smaller hydrophobic residues such as Gly, Ala and Val to line the pocket of Ma-WS/DGAT (Figure 2. 6 D-F). In addition to these amino acid substitutions, the relative disposition of helix α4 and the β-sheet of the N-terminal domain is different among these acyltransferases and also results in the formation of differently sized binding pocket (Figure 2. 7).

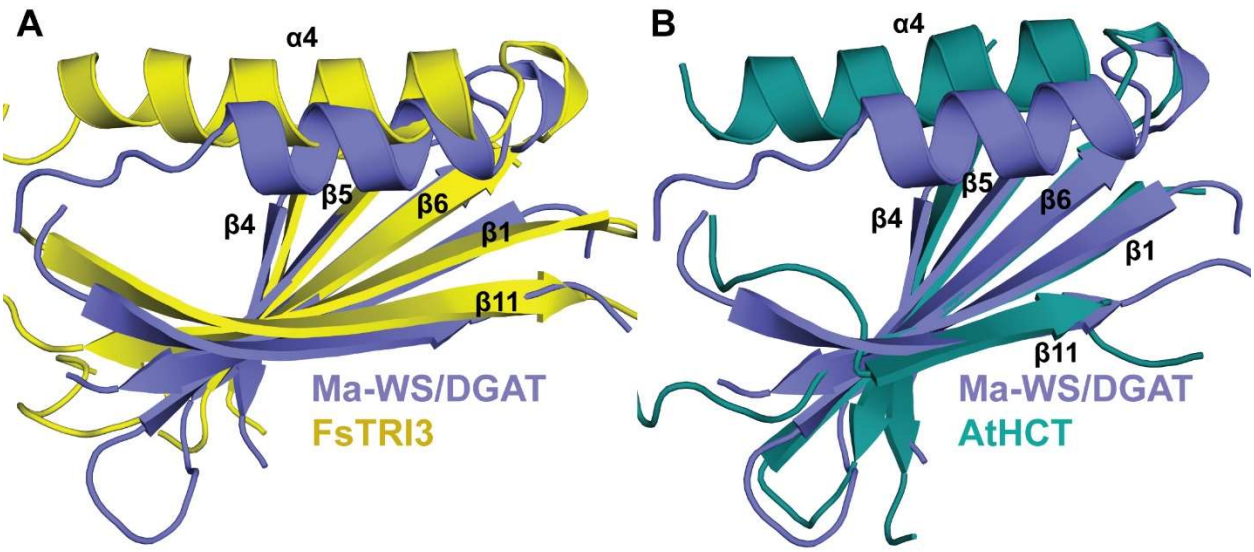

Figure 2. 7: Superposition of WS/DGAT from *Marinobacter aquaeolei* VT8 (Ma-WS/DGAT) with (A) TRI3 trichothecene 15-O-acetyltransferase from *Fusarium sporotrichioides* (FsTRI3)[2] and (B) hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase from *Arabidopsis thaliana* (AtHCT)[3]. The figure shows the difference in the relative disposition of helix α4 and the β-sheet of the N-terminal domain in these acyltransferases. Ma-WS/DGAT is shown in blue, FsTRI3 in yellow and AtHCT in teal. The numbering of the secondary structure is based on the structure of Ma-WS/DGAT.

To further investigate whether pocket 1 encompasses the acyl chain of palmitoyl-CoA, we generated site-specific variants and compared their activity with that of the wild type Ma-WS/DGAT (Table 2. 4). We selected Gly25 and Ala144 to be mutated because of their position in pocket 1 and their average sequence conservation (Figure 2. 8A). Specifically, we generated Gly25→Val and Ala144→Val/Phe variants in an attempt to "shorten" pocket 1 so that acyl-CoAs more than 7 and 6 carbons long respectively do not "fit" if bound in a fully extended conformation. These variants were tested for activity using palmitoyl-CoA (C16-CoA) and their activities were compared against wild type (WT) Ma-WS/DGAT. The Gly25→Val mutation had an immense effect on the activity of the enzyme towards palmitoyl-CoA (1.0% of WT) while the Ala144→Val/Phe variants were still ~ 40-50% active (Table 2. 4). Although, binding of palmitoyl-CoA was not expected in the case of the Ala144→Val/Phe mutants, these results suggest that

Table 2. 4: Activity of wild type Ma-WS/DGAT and mutants. The activity of the mutants was expressed as percentage of WE formation in respect to the wild type reaction (considered as 100%). Results are means of duplicate experiments. (ND = Not Determined)

Ma-WS/DGAT	hexyl-palmitate (%)	hexyl-caproate (%)
WT	100	100
H136N	1.5	ND
D140A	49.6	ND
D8A	24.1	ND
G25V	1.0	46.9
A144V	43.3	ND
A144F	53	654.6

palmitoyl-CoA may still bind in a different, less favorable conformation where the acyl chain is folded instead of fully extended. This binding mode may energetically be even less favorable in the case of the Gly25→Val mutant where pocket 1 is expected to be long enough to fit a 7-carbon long chain instead of 6. In addition to palmitoyl-CoA, we tested the activity of the Gly25→Val and Ala144→Phe variants towards hexanoyl-CoA (C6-CoA). The Gly25→Val variant was 46.9% active compared to WT Ma-WS/DGAT while a 5.5-fold increase in activity was observed for the Ala144→Phe mutant (Table 2. 4). This increased activity of the mutants towards the shorter acyl-CoA further supports the role of pocket 1 as the acyl chain-binding site.

To compare the substrate preference of wild type, Gly25→Val and Ala144→Phe Ma-WS/DGATs for acyl-CoA, we performed competition assays. Briefly, both hexanoyl- and palmitoyl-CoA were added in the reactions at equal concentrations, and the wax esters formed (hexyl-caproate and hexyl-palmitate respectively) were expressed as percentage of total wax esters

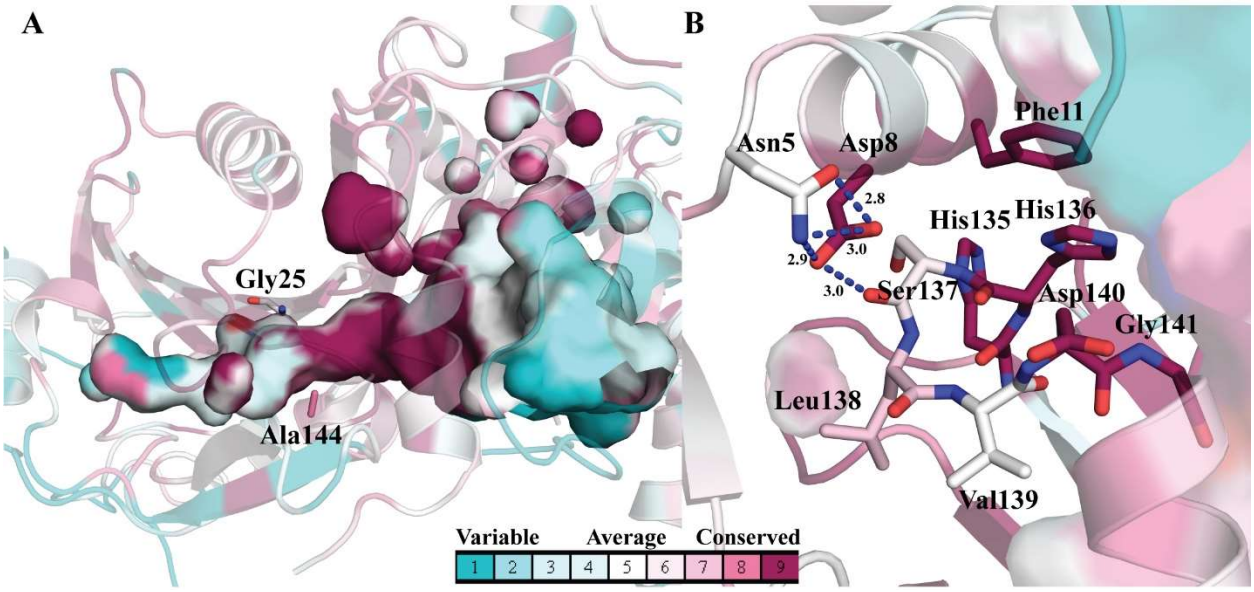

Figure 2. 8: Ribbon diagram and cavity of Ma-WS/DGAT colored based on amino acid conservation as calculated by ConSurf. (A) Gly25 and Ala144 are located on strand β1 and helix α4 respectively lining the cavity where the acyl chain binds. (B) Positioning of His136 for catalysis by the conserved motif loop (shown in sticks) and Phe11. Asp8 is also involved in maintaining the structural integrity of the loop. Distances shown are in Å.

produced by the enzyme. Both mutants showed a preference for the shorter acyl-CoA with Gly25→Val producing almost exclusively hexyl-caproate (~90% of total wax esters formed), and Ala144→Phe producing ~65% hexyl-caproate and ~35% hexyl-palmitate (Figure 2. 9). Although the Gly25→Val variant showed a clear preference for hexanoyl-CoA instead of palmitoyl-CoA, the Ala144→Phe was more efficient as it produced more of both products. Nonetheless, these results further demonstrate that the acyl chain of the acyl-CoA binds in pocket 1 as well as that Ma-WS/DGAT is amenable to structure-guided engineering towards the production of desired wax esters.

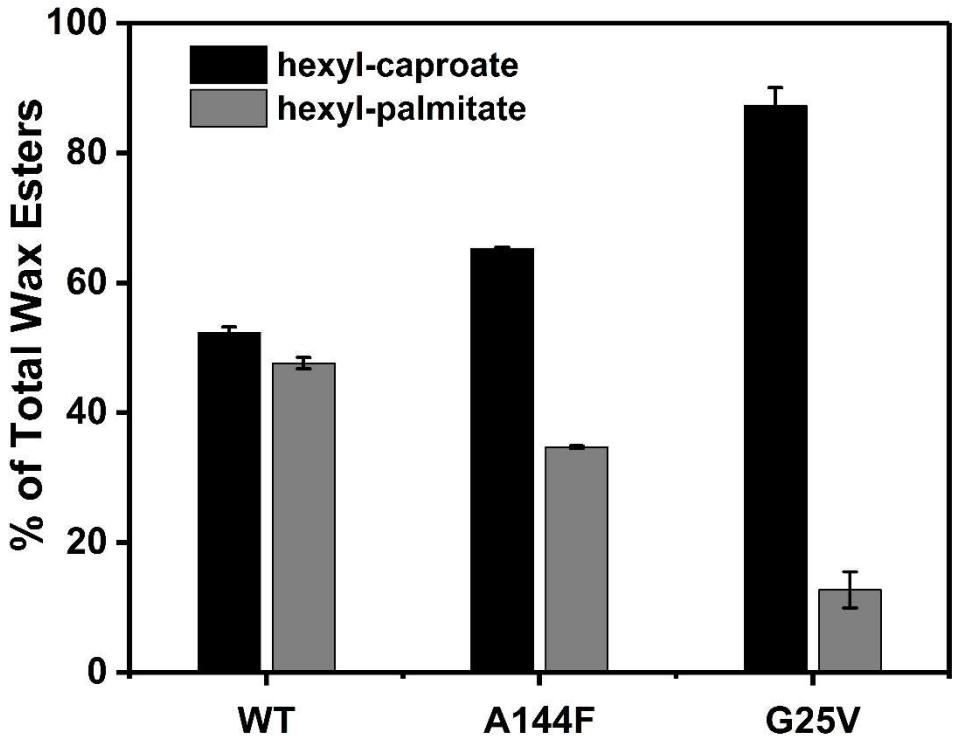

Figure 2. 9: Percentage of total wax esters produced by Ma-WS/DGAT wild type (WT) and mutants when hexanol and both hexanoyl- and palmitoyl-CoA were added in the reactions. Results are means ± SEM of duplicate experiments.

Role of Conserved Motif Residues

The proposed mechanism for WS/DGATs requires abstraction of a proton from an alcohol by a His residue to generate the reactive alkoxide (Figure 2. 1). However, the side chain of His is not sufficiently nucleophilic to abstract a proton from an alcohol (pK_a of about 17). This raises the question how the catalytic His can abstract hydroxyl proton from the acyl acceptor in the active site of WS/DGATs. In the crystal structure of Ma-WS/DGAT, His136 is positioned for catalysis in two ways. First, the other residues of the conserved motif HHXXXDG (HHSLVDG in Ma-WS/DGAT) form interactions that maintain the structural integrity of the loop and position His136 towards the enzyme cavity (Figure 2. 3C-D). Second, the imidazole ring of His136 is oriented by π interactions with the aromatic ring of Phe11 (Figure 2. 8B). There are no residues, however, within hydrogen bonding distance from His136. The only other highly conserved residue adjacent

to the catalytic His is Asp8 (Figure 2. 8B). This led us to the hypothesis that upon binding of the substrates, Asp8 might move closer to His136 and tune its pK_a. To test this hypothesis, we generated an Asp8→Ala mutant and found that its activity was about half of the wild type (Table 2. 4). However, a closer examination of the structure revealed that it is hydrogen bonded to the carbonyl oxygen of Ser137 (Figure 2. 8B). This suggests that Asp8 may play a structural role in proper positioning of His136 for catalysis by participating in maintaining the structural integrity of the loop formed by the conserved motif. In the structure of AtHCT, the catalytic His153 is in hydrogen bonding distance from Asn300. Comparatively, the structure of Ma-WS/DGAT is missing residues 296-308 which include the highly conserved Asn306 and semi-conserved Glu307 (Figure 2. 10). It is likely that these residues tune the pK_a of His136 or that the environment of the active site makes His136 a stronger base for catalysis.

Figure 2. 10: Sequence logo generated by WebLogo (http://weblogo.berkeley.edu/) using as input the curated seed alignment of PFAM for wax synthases (PF03007) [57-59]. Numbering of residues corresponds to the sequence of Ma-WS/DGAT.

2.4 Concluding Remarks

Our structural analysis of Ma-WS/DGAT reveals an acyltransferase-like fold that has been adapted to accommodate large hydrophobic substrates. The structural data in conjunction with our mutational analysis suggest that the acyl chain of the acyl-CoA binds in a pocket formed at the

core of the N-terminal domain. Notably, introduction in this pocket of the highly hydrophobic Phe at position 144 resulted in an enzyme that was 6.5 times more active towards a shorter acyl-CoA (C6-CoA) compared to the wild type. These results, in combination with previous studies [24], further support the notion that the bacterial WS/DGATs are amenable to site-directed engineering. This work provides a framework for further engineering studies aimed at *in vivo* production of desired wax esters utilizing bacterial WS/DGATs.

2.5 References

1. Vanhercke, T., et al., *Metabolic engineering of plant oils and waxes for use as industrial feedstocks.* Plant Biotechnol J, 2013. **11**(2): p. 197-210.

2. Commission, I.W. *Catch Limits & Catches taken*. [cited 2017 October 23]; Available from: https://iwc.int/catches.

3. Singh, S., A. Bhadani, and B. Singh, *Synthesis of Wax Esters from α-Olefins.* Industrial & Engineering Chemistry Research, 2007. **46**(8): p. 2672-2676.

4. Deng, L., et al., *Synthesis of Wax Esters by Lipase-catalyzed Esterification with Immobilized Lipase from Candida sp. 99–125.* Chinese Journal of Chemical Engineering, 2011. **19**(6): p. 978-982.

5. Santala, S., et al., *Rewiring the wax ester production pathway of Acinetobacter baylyi ADP1.* ACS Synthetic Biology, 2014. **3**(3): p. 145-51.

6. Kuo, C.-H., et al., *High Yield of Wax Ester Synthesized from Cetyl Alcohol and Octanoic Acid by Lipozyme RMIM and Novozym 435.* International Journal of Molecular Sciences, 2012. **13**(9): p. 11694-11704.

7. Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.* Nature, 2010. **463**(7280): p. 559-62.

8. Kalscheuer, R., T. Stolting, and A. Steinbuchel, *Microdiesel: Escherichia coli engineered for fuel production.* Microbiology, 2006. **152**(Pt 9): p. 2529-36.

9. Biester, E.M., et al., *Identification of avian wax synthases.* BMC Biochem, 2012. **13**: p. 4.

10. Biester, E.M., J. Hellenbrand, and M. Frentzen, *Multifunctional acyltransferases from Tetrahymena thermophila.* Lipids, 2012. **47**(4): p. 371-81.

11. Teerawanichpan, P. and X. Qiu, *Fatty acyl-CoA reductase and wax synthase from Euglena gracilis in the biosynthesis of medium-chain wax esters.* Lipids, 2010. **45**(3): p. 263-73.

12. Kalscheuer, R. and A. Steinbuchel, *A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in Acinetobacter calcoaceticus ADP1.* J Biol Chem, 2003. **278**(10): p. 8075-82.

13. Rottig, A. and A. Steinbuchel, *Acyltransferases in bacteria.* Microbiol Mol Biol Rev, 2013. **77**(2): p. 277-321.

14. Villa, J.A., et al., *Use of limited proteolysis and mutagenesis to identify folding domains and sequence motifs critical for wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase activity.* Appl Environ Microbiol, 2014. **80**(3): p. 1132-41.

15. Turchetto-Zolet, A.C., et al., *Diversity and evolution of plant diacylglycerol acyltransferase (DGATs) unveiled by phylogenetic, gene structure and expression analyses.* Genet Mol Biol, 2016. **39**(4): p. 524-538.

16. King, A., et al., *Cuticular wax biosynthesis in petunia petals: cloning and characterization of an alcohol-acyltransferase that synthesizes wax-esters.* Planta, 2007. **226**(2): p. 381-94.

17. Li, F., et al., *Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis.* Plant Physiol, 2008. **148**(1): p. 97-107.

18. Zhang, N., et al., *Two bifunctional enzymes from the marine protist Thraustochytrium roseum: biochemical characterization of wax ester synthase/acyl-CoA:diacylglycerol acyltransferase activity catalyzing wax ester and triacylglycerol synthesis.* Biotechnol Biofuels, 2017. **10**: p. 185.

19. Barney, B.M., et al., *Differences in substrate specificities of five bacterial wax ester synthases.* Appl Environ Microbiol, 2012. **78**(16): p. 5734-45.

20. Holtzapple, E. and C. Schmidt-Dannert, *Biosynthesis of isoprenoid wax ester in Marinobacter hydrocarbonoclasticus DSM 8798: identification and characterization of isoprenoid coenzyme A synthetase and wax ester synthases.* J Bacteriol, 2007. **189**(10): p. 3804-12.

21. Kalscheuer, R., et al., *Neutral lipid biosynthesis in engineered Escherichia coli: jojoba oil-like wax esters and fatty acid butyl esters.* Appl Environ Microbiol, 2006. **72**(2): p. 1373-9.

22. Schmidt-Dannert C, H.E., *Isoprenoid wax ester synthases, isoprenoid acyl CoA-synthetases, and uses thereof.* . 2009.

23. Jens Nielsen, S.S., *Methods and products for production of wax esters (u.s. national phase)* 2013.

24. Barney, B.M., et al., *Altering small and medium alcohol selectivity in the wax ester synthase.* Appl Microbiol Biotechnol, 2015. **99**(22): p. 9675-84.

25. Uthoff, S., et al., *Thio wax ester biosynthesis utilizing the unspecific bifunctional wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase of Acinetobacter sp. strain ADP1.* Appl Environ Microbiol, 2005. **71**(2): p. 790-6.

26. Kalscheuer, R., H. Luftmann, and A. Steinbuchel, *Synthesis of novel lipids in Saccharomyces cerevisiae by heterologous expression of an unspecific bacterial acyltransferase.* Appl Environ Microbiol, 2004. **70**(12): p. 7119-25.

27. Shi, S., et al., *Functional expression and characterization of five wax ester synthases in Saccharomyces cerevisiae and their utility for biodiesel production.* Biotechnol Biofuels, 2012. **5**: p. 7.

28. Teo, W.S., et al., *Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid short- and branched-chain alkyl esters biodiesel.* Biotechnol Biofuels, 2015. **8**: p. 177.

29. Zhou, Y.J., et al., *Fatty Acid-Derived Biofuels and Chemicals Production in Saccharomyces cerevisiae.* Front Bioeng Biotechnol, 2014. **2**: p. 32.

30. Santala, S., et al., *Rewiring the Wax Ester Production Pathway of Acinetobacter baylyi ADP1.* ACS Synthetic Biology, 2014. **3**(3): p. 145-151.

31. Wierzbicki, M., et al., *Engineering an Escherichia coli platform to synthesize designer biodiesels.* J Biotechnol, 2016. **224**: p. 27-34.

32. Rottig, A., P.J. Zurek, and A. Steinbuchel, *Assessment of bacterial acyltransferases for an efficient lipid production in metabolically engineered strains of E. coli.* Metab Eng, 2015. **32**: p. 195-206.

33. De Crecy-Lagard, V., P. Marliere, and W. Saurin, *Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation.* C R Acad Sci III, 1995. **318**(9): p. 927-36.

34. Buglino, J., et al., *Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from Mycobacterium tuberculosis.* J Biol Chem, 2004. **279**(29): p. 30634-42.

35. Kleanthous, C., P.M. Cullis, and W.V. Shaw, *3-(Bromoacetyl)chloramphenicol, an active site directed inhibitor for chloramphenicol acetyltransferase.* Biochemistry, 1985. **24**(20): p. 5307-13.

36. Leslie, A.G., *Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 A resolution.* J Mol Biol, 1990. **213**(1): p. 167-86.

37. Stöveken, T., R. Kalscheuer, and A. Steinbüchel, *Both histidine residues of the conserved HHXXXDG motif are essential for wax ester synthase/acyl-CoA:diacylglycerol acyltransferase catalysis.* European Journal of Lipid Science and Technology, 2009. **111**(2): p. 112-119.

38. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.

39. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-26.

40. Vonrhein, C., et al., *Data processing and analysis with the autoPROC toolbox.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 293-302.

41. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(Pt 2): p. 133-144.

42. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.

43. Grosse-Kunstleve, R.W. and P.D. Adams, *Substructure search procedures for macromolecular structures.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 1966-73.

44. McCoy, A.J., L.C. Storoni, and R.J. Read, *Simple algorithm for a maximum-likelihood SAD function.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 7): p. 1220-8.

45. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(2): p. 213-221.

46. Terwilliger, T.C., et al., *Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 6): p. 582-601.

47. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallographica Section D: Biological Crystallography, 2008. **64**(Pt 1): p. 61-69.

48. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D, Biological Crystallography, 2006. **62**(Pt 9): p. 1002-11.

49. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallographica Section D, Biological Crystallography, 2008. **64**(Pt 1): p. 83-9.

50. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 235-42.

51. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(4): p. 486-501.

52. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 355-67.

53. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.

54. Read, R.J., et al., *A new generation of crystallographic validation tools for the protein data bank.* Structure, 2011. **19**(10): p. 1395-412.

55. Landau, M., et al., *ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W299-302.

56. Ashkenazy, H., et al., *ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.* Nucleic Acids Res, 2016. **44**(W1): p. W344-50.

57. Schneider, T.D. and R.M. Stephens, *Sequence logos: a new way to display consensus sequences.* Nucleic Acids Res, 1990. **18**(20): p. 6097-100.

58. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.

59. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future.* 2016. **44**(D1): p. D279-85.

60. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Research, 2010. **38**(Web Server issue): p. W545-9.

61. Garvey, G.S., et al., *Structural and functional characterization of TRI3 trichothecene 15-O-acetyltransferase from Fusarium sporotrichioides.* Protein Sci, 2009. **18**(4): p. 747-61.

62. Levsh, O., et al., *Dynamic Conformational States Dictate Selectivity toward the Native Substrate in a Substrate-Permissive Acyltransferase.* Biochemistry, 2016. **55**(45): p. 6314-6326.

CHAPTER 3: STRUCTURE-FUNCTION STUDIES OF THE LOGANIC ACID *O*-METHYLTRANSFERASE FROM *Catharanthus roseus*[2]

3.1 Introduction

Monoterpenoid indole alkaloids (MIAs) are a group of plant natural products with a number of medicinal uses including chemotherapeutics (Vinca alkaloids), antihypertensives (ajmalicine and serpentine), and antidiabetics (vindolicine) [1-3]. These nitrogen-containing compounds are produced by the Mannich-type condensation of tryptamine with secologanin to yield strictosidine, which serves as the common intermediate for the ~3000 different MIAs of diverse chemical structures [4]. Currently, isolation of MIAs is dependent on direct extraction from plants, and the dispersion of the biosynthetic steps across different cells types, as well as the trace production levels *in situ* undermines cost-effective isolation of these molecules, and restricts detailed exploration of MIA pharmacognosy [5, 6]. The elucidation of the complete nine-step biosynthetic pathway for the production of secologanin from the primary metabolite geranyl pyrophosphate in the model plant *Catharanthus roseus* (Madagascar periwinkle) [7] has prompted efforts towards the heterologous production of the Vinca alkaloid vindoline [8], as well as the common precursor strictosidine in both Baker's yeast (*Saccharomyces cerevisiae*) [9] and in a plant host (*Nicotiana benthamiana*) [7]. The last step in this seco-iridoid pathway involved the conversion of the

[2] This chapter is adapted from the following published article:

Nektaria Petronikolou, Allison Hollatz, Mary A. Schuler & Satish K. Nair (2018). Loganic Acid Methyltransferase: Insights into the specificity of methylation on an iridoid glycoside. *Chembiochem*, doi: 10.1002/cbic.201700679.

cyclopentanopyran monoterpenoid glycoside loganic acid into loganin [10]. Loganin is an iridoid glycoside that itself has recently attracted interest as a drug candidate due to its reported neuroprotective properties, cognitive-inducing activity, and inhibitory effect on melanogenesis among other biological activities [11-15].

The conversion of loganic acid into loganin is catalyzed by loganic acid methyltransferase (LAMT), via the transfer of a methyl group from S-adenosyl-methionine (SAM) to the carboxylate of loganic acid (LA) [10, 16]. Random sequencing of *C. roseus* leaf epidermis-enriched cDNA library identified multiple genes encoding for putative SAM-dependent *O*-methyltransferases, and biochemical characterization of a heterologously expressed candidate LAMT from *C. roseus* (CraLAMT) confirmed its function as a loganic acid methyltransferase [10]. Bioinformatics analysis suggests that CraLAMT is a member of the SABATH family of methyltransferases [10], which include enzymes that modify small secondary metabolites such as plant hormones and signaling molecules. The biological activity of such plant metabolites is regulated by the addition or removal of the methyl group, implicating the network of SABATH enzymes in numerous processes that regulate plant responses to stress [17].

In contrast to other characterized SABATH methyltransferases that demonstrate broad substrate tolerance [18], CraLAMT has been shown to be remarkably specific for only its physiological iridoid substrate [10]. Specifically, CraLAMT only methylates the glycosylated and hydroxylated iridoid ring ring rather than the aglycone (loganetic acid) or deoxyloganic acid [10, 16]. The enzyme also demonstrates strict stereospecificity for loganic acid (6S,7R) as a substrate and fails to methylate either 7-epiloganic acid (6S,7S), or the oxidized 7-keto dehydrologanic acid.

In addition to the strict substrate specificity, CrLAMT is further distinguished by the unique nature of its glycosylated substrate. Canonical plant SABATH methyltransferases function on

bulky hydrophobic compounds, such as salicylic acid, benzoic acid and indole-3-acetic acid. In contrast, the loganic acid substrate of CrLAMT contains a hydrophilic glucose moiety and several additional polar groups. In order to identify the molecular rationale for its unique substrate specificity that distinguishes it from canonical SABATH methyltransferases, we carried out biochemical characterizations and crystallographic structure determination of CrLAMT in complex with loganic acid, and mutational analyses of active site residues involved in catalysis.

3.2 Materials and Methods

Chemicals and Reagents

All chemical reagents were purchased from Sigma-Aldrich unless otherwise noted.

Cloning and Site-Specific Mutagenesis

Coding sequences for the full-length loganic acid methyltransferase from *Catharanthus roseus* (CraLAMT) (Genbank EU057974) were initially reverse transcribed by PCR from *Catharanthus roseus* leaf RNA and inserted into the pGEM-T easy vector (Promega). Among several sequence variants obtained, the one identical to Genbank EU057974 was PCR-amplified with CraLAMT_WT_F and CraLAMT_WT_R (Table B. 1), and inserted into NcoI-XhoI cut pET28a (Invitrogen) in-frame with the C-terminal His_6-tag in this expression vector. After confirmation of the full-length CraLAMT sequence, site-directed mutations were introduced using a PCR-based site-directed mutagenesis strategy with each reaction consisting of 25 ng of the His_6-tagged CraLAMT template in the pET28a cloning vector, 1.0 μM of each forward and reverse site-directed mutagenesis primer (Table B. 1), 200 μM each dNTP, 2.5 units Pfu Ultra HF DNA polymerase (Agilent), 1X Pfu Ultra HF Buffer (Agilent), and molecular biology grade water

(Corning) to bring the total reaction volume to 25 µl. The reactions began with a 3 min denaturation step at 95 °C, proceeded with 17 cycles of 60 sec denaturation at 95 °C, 60 sec annealing at 36 - 44 °C, 12 min extension at 68 °C and were followed with an 8 min extension at 68 °C and multiple hours at 4 °C. These PCR products were treated with 1 µl DpnI (Fermentas) in 1X CutSmart buffer (NEB) overnight at 37 °C to digest the original methylated His_6-tagged CraLAMT template and the DpnI-digested PCR products were transformed into BL21 cells and screened by sequence analyses. All site-directed mutants were sequenced in their entirety to confirm the absence of other mutations.

Protein Expression and Purification

For protein expressions, all CraLAMT expression constructs were transformed into *E. coli* strain BL21 (Invitrogen) and inoculated into two 1.25 L cultures containing 50 µg/ml kanamycin. Cultures were incubated at 37 °C for 4 - 6 hours with shaking (180 rpm) until the OD_{600} readings reached 0.50 - 0.75, induced with 0.143 g/L IPTG and incubated at 25 °C for an additional 24 hours. For protein purifications, cell pellets from the 2.5 L cultures were resuspended in 60 ml of Buffer A (50 mM NaP_i - pH 8.0, 0.3 M NaCl, 10 mM imidazole), treated for one hour at 25 °C with 1 mg/ml lysozyme, sonicated at least three times for 15 sec and centrifuged at 17,000 rpm for 30 min in a SS34 rotor (Sorvall). Each cleared supernatant was loaded onto a 10 ml Ni-NTA agarose affinity column (Qiagen) pre-equilibrated with Buffer A, washed with an additional 30 ml of Buffer A and 50 ml of Buffer B (50 mM NaP_i - pH 8.0, 0.3 M NaCl, 50 mM imidazole). Bound proteins were eluted with 17 ml of Buffer C (50 mM NaP_i - pH 8.0, 0.3 M NaCl, 250 mM imidazole), and dialyzed at 4 °C against four changes of 375 ml of Buffer D (50 mM Tris-HCl, pH 7.5, 200 mM NaCl) (1.5 L total). Dialyzed protein samples were centrifuged for 1 min at 6000 rpm

in a clinical centrifuge to remove precipitated protein, adjusted to 10% glycerol and flash frozen in liquid nitrogen. No glycerol was added in the protein samples used for crystallization. Protein concentrations were determined by NanoDrop (Thermo Fisher Scientific) analyses using the molecular weights and extinction coefficients calculated by the ProtParam tool (ExPASY server) (Table B. 2) [19].

Crystallization

Initial crystallization conditions were determined by the sparse matrix sampling method using commercial screens. Crystals of the CraLAMT-SAH (S-adenosyl-homocysteine) complex were grown using the sitting drop vapor diffusion method. Briefly, 1 µl of protein at 5 mg/ml concentration was incubated with 1 mM SAH (Sigma-Aldrich) for 30 min on ice, and was subsequently mixed with 1 µl of precipitant solution (20% PEG 6000, 0.1 M HEPES-NaOH, pH 7.0, 0.2 M NaCl), and equilibrated with a well containing the precipitant solution at 9 °C. Crystals grew within few days, and were briefly soaked in precipitant solution supplemented with 30% ethylene glycol prior to flash-cooling in liquid nitrogen.

Crystals of the CraLAMT-SAH-LA complex were grown using the hanging drop vapor diffusion method: 1 µl of protein at 5 mg/ml concentration was incubated with 4 mM LA (Sigma-Aldrich) for 1 h at room temperature, and was subsequently mixed with 1 µl of precipitant solution (18% PEG 6000, 0.1 M MES-NaOH, pH 7.0, 0.2 M NaCl), and equilibrated over a well containing the precipitant solution at 9 °C. Once crystals were formed, they were transferred into a drop containing the precipitant solution and 4 mM LA for 1 h at 9 °C, and were subsequently transferred into a drop containing the precipitant solution, 4 mM LA and 2 mM SAH for 1 h at 9 °C. Crystals

were briefly soaked in precipitant solution supplemented with 30% ethylene glycol, 5 mM LA and 2 mM SAH prior to flash-cooling in liquid nitrogen.

Data Collection, Phasing and Structure Determination

X-ray diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT), Sector 21, Argonne National Laboratory. All data were indexed, integrated and scaled using either HKL2000 or AutoProc [20-23]. Both structures were determined by molecular replacement as implemented in the Phenix program suite [24]; for the CraLAMT-SAH structure, the coordinates of salicylic acid methyltransferase [25] (PDB ID: 1M6E) were used as search model, while for the CraLAMT-SAH-LA complex the refined coordinates of CraLAMT-SAH were used. In both cases, the resultant solutions were subsequently used as starting models for several rounds of automated model building using either the Phenix Autobuild wizard [26] or Buccanneer [27, 28] as implemented in the CCP4 suite of programs [29], followed by rounds of manual rebuilding using COOT [30], combined with crystallographic refinement using REFMAC5 [31]. Ligands were built in COOT, and water molecules were added using the ARP/wARP solvent building software of the CCP4 suite [32], and confirmed by manual inspection. The quality of the in-progress model was routinely monitored using both the free R factor [33] and MolProbity [34] for quality assurance. Data collection and refinement statistics are presented in Table 3. 1. For the CraLAMT-SAH-LA complex, clear electron density for loganic acid was only observed in one monomer. Although sparse density was observed in the other monomer, it was not adequate for the ligand to be built with confidence.

Table 3. 1: Data collection and refinement statistics. (1. Highest resolution shell is shown in parenthesis, 2. $R_{sym} = \Sigma |(I_i - \langle I_i \rangle | \Sigma I_i$ where I_i = intensity of the *i*th reflection and $\langle I_i \rangle$ = mean intensity, 3. R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.)

	SAH	**SAH-LA**
Data collection		
Space Group	$P2_12_12_1$	$P2_12_12_1$
a, b, c (Å)	57.8, 116.2, 136.3	58.4, 116.2, 136.6
Resolution (Å)	35.0-2.2	116.2-1.95
R_{sym} (%)[1,2]	12.8 (83.6)	10.9 (103.8)
I/σ(I)	15.9 (2.0)	15.5 (2.1)
Completeness (%)	100 (99.5)	99.6 (96.5)
Redundancy	8.0 (7.4)	7.4 (7.5)
Total reflections	50,471	504,320
Unique reflections	45,507	68,324
Refinement		
Resolution (Å)	34.1-2.2	68.3-1.95
No. reflections used	45,501	68,223
R_{work} / R_{free}[3]	18.5/22.9	17.5/21.9
Number of atoms		
Protein	5687	5781
Loganic acid	-	26
SAH	52	52
Water	280	7670
B-factors		
Protein	35.9	30.8
Loganic acid	-	27.3
SAH	23.9	20.1
Water	39.0	39.6
R.m.s deviations		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	0.917	1.073
MOLPROBITY statistics		
Clash score	5.38	2.42
Ramachandran outliers/ allowed/ favored (%)	0.00/ 2.84/ 97.16	0.00/ 2.92/ 97.08

Determination of Kinetic Parameters

The kinetic parameters of the wild type protein were determined by detecting loganin formation by hydrophilic interaction liquid chromatography (HILIC). All enzyme reactions were performed in 20 mM Tris-HCl (pH 7.5) and 100 mM NaCl at 30 °C. For determination of the Michaelis-Menten parameters of the wild type enzyme for loganic acid, a 150 µl reaction volume contained the following components: 0.5 µM CraLAMT, 5 µM SAH nucleosidase, 400 µM SAM (Sigma-Aldrich), and various concentrations of loganic acid (Cambridge Chemicals). The reactions were stopped by addition of 300 µl acetonitrile (ACN), flash-frozen in liquid nitrogen, and dried for 1-1.5 h with a vacuum concentrator (Labconco CentriVap). The dried material was resuspended into 150 µl methanol/ACN (25/75 v/v), and the precipitated protein was removed by centrifuging for 4 min at 4 °C. 100 µl of the reaction mixture was then injected into a HPLC (Shimadzu prominence LC 20A) equipped with a Diol-HILIC column (YMC-Triart, 5 µm particle size, 4.6 x 250 mm). Solvent A consisted of 90% ACN and 10% of 25 mM Ammonium Acetate-pH 5.3, and solvent B consisted of 70 % ACN and 30% of 25 mM Ammonium Acetate-pH 5.3 (pH was adjusted prior to addition of the organic solvent). A gradient elution was applied as follows: 10 min of 10% B, 2 min gradient from 10% to 100% B, 25 min of 100% B, 2 min gradient from 100% back to 10% B and 20 min of 10% B. The flow rate was 1 ml/min throughout the procedure, and the separation of the reaction components was monitored at 240 nm.

The area under the loganin peak was converted to concentration using a loganin calibration curve (Figure B. 1) generated with known concentrations of standard, using the same column and instrument. Control reactions without addition of the enzyme were also included and subtracted from the enzyme reactions. Based on the initial reaction rates, the apparent K_M and V_{max} values

were determined using the Michaelis - Menten function of Origin (OriginLab, Northampton, MA). Results are means ± SEM of triplicate experiments.

The commercial SAM used in these experiments was further purified and quantified as previously described [35]. The stability and integrity of each reaction component after drying the samples with the vacuum concentrator was assessed by HILIC analysis of standards before and after drying them with the vacuum concentrator. All components were found to be stable during this process except for SAM which was converted to methylthioadenosine (MTA). The MTA peak was separated from the loganin peak (Table 3. 2) and did not affect our analysis. The identity of each peak was confirmed by high resolution liquid chromatography-mass spectrometry (LC/MS) (Table 3. 3).

Table 3. 2: Retention times of each reaction component as analyzed by hydrophilic interaction liquid chromatography (HILIC).

Compound	Retention Time
Methylthioadenosine	5.8
Loganin	6.3
Adenine	9.2
Loganic acid	24.1
SAH	25.2
SAM	34.7

End-Point Activity Assay for CraLAMT Wild Type and Mutants

For the activity of the wild type CraLAMT (Figure 3. 1B), 100 µl reactions containing 10 µM CraLAMT, 5.5 µM Pfs, 1.2 mM SAM and 600 µM loganic acid were incubated at 30 °C for 30 min, stopped by addition of 200 ul of ACN, vortexed, span down and analyzed by HILIC as described above (without drying the samples with the vacuum concentrator).

In order to compare the activity of the wild type CraLAMT and the mutants, reactions were analyzed by HILIC as described above. 150 µl reactions containing 0.5 µM enzyme, 5.5 µM SAH nucleosidase, 1.2 mM SAM and 600 µM (~ $2K_M$) loganic acid were incubated at 30 °C for 30 min. The reactions were stopped, treated and analyzed identically to the reactions for kinetic analysis. The activity of the mutants was expressed as percentage of loganin formation in respect to the wild type reaction (considered as 100%). For this purpose, the area under the loganin peak of the wild type reaction was considered as 100%, and the areas under the loganin peak of the mutants were converted to % of activity with the following equation: % of activity = (peak area of mutant) x 100 / (peak area of wild type). Control reactions without addition of the enzyme were also analyzed. Results are means ± SEM of triplicate experiments.

High Resolution Liquid Chromatography-Mass Spectrometry (LC/MS)

For high resolution LC/MS, the samples were analyzed by using the Q-Exactive MS system (Thermo. Bremen, Germany) in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Software Xcalibur 3.0.63 was used for data acquisition and analysis. The Dionex Ultimate 3000 series HPLC system (Thermo, Germering, Germany) used includes a DAD, a degasser, an autosampler, and a binary pump. The LC separation was performed with the same column, solvents and method used for the kinetic analysis. The autosampler was set to 15 °C. The injection volume was 40 µL. Mass spectra were acquired under both negative and positive electrospary ionization (ESI) (sheath gas flow rate, 65; aux gas flow rate: 20; sweep gas flow rate, 4; spray voltage, 3.5 kV; capillary temp, 300 °C; aux gas heater temperature, 500 °C. The resolution was set to 70,000. The AGC target was 1E6 with a maximum injection time of 200 ms.

Table 3. 3: High resolution LC/MS analysis of the reactions (10 uM CraLAMT, 5.5 uM Pfs, 1.2 mM SAM, 600 uM loganic acid - 30 min at 30 °C) to further confirm the identity of each peak.

Compound	m/z
Methylthioadenosine	298.096 $[M+H]^+$
Loganin	391.159 $[M+H]^+$
Adenine	136.062 $[M+H]^+$
Loganic acid	375.130 $[M-H]^-$
SAM	399.144 $[M+H]^+$

3.3 Results and Discussion

Using purified recombinant CraLAMT, we first confirmed that the enzyme was active (Figure 3. 1B), and determined the Michaelis - Menten kinetics parameters for the enzyme using loganic acid as the substrate (Figure 3. 1C). Because most SAM-dependent methyltransferases are known to be inhibited by product S-adenosyl-homocysteine (SAH), SAH nucleosidase was included in all of the assays (the Pfs SAH nucleosidase has a catalytic efficiency of 11.6×10^6 $M^{-1}s^{-1}$) (Figure 3. 1B) [36]. The catalytic efficiency of CraLAMT (k_{cat}/K_M) was determined to be 996 $M^{-1}s^{-1}$ (pH 7.5, 30 °C) (Figure 3. 1C), with a relatively high K_M (315 ± 35 μM) compared to canonical plant SABATH O-methyltransferases [e.g. 23 ± 6.3 μM for the salicylic acid *O*-methyltransferase from *Clarkia brewe*ri (CbSAMT) and 13 ± 5.1 μM for the indole-3-acetic acid *O*-methyltransferase from *Arabidopsis thaliana* (AtIAMT)] [25]. In a previous study [10], the K_M of CraLAMT was reported to be about 50 times higher than the one reported here (14,700 ± 1,700 μM at pH 7.5, 37 °C). However, our assays were performed using enzyme of high purity (>95%) instead of crude preparations [10], and SAH nucleosidase was used in our assays in order to avoid product inhibition.

In order to identify features that dictate accommodation of the glycosylated substrate, we determined the 1.95 Å resolution co-crystal structure of CraLAMT in complex with SAH and loganic acid. The core architecture of CraLAMT recapitulates the general features of other SABATH methyltransferases (Figure B. 2), such as the CbSAMT (PDB ID 1M6E) [25], the AtIAMT (PDB ID 3B5I) [37] and the xanthosine *N*-methyltransferase from *Coffea canephora* 'robusta' (CcXMT) (PDB ID 2EG5) [38]. Specifically, the overall fold consists of a Rossmann-

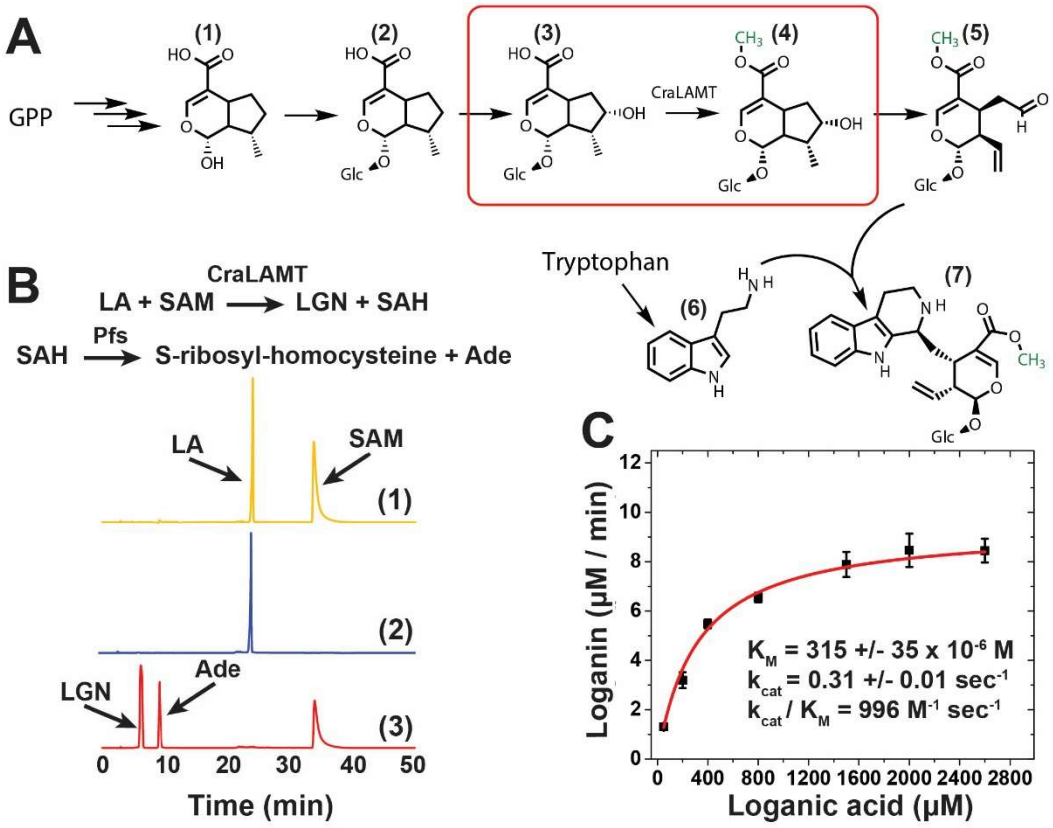

Figure 3. 1: (A) Biosynthesis of strictosidine in Catharanthus roseus. Strictosidine is the precursor of many monoterpenoid indole alkaloids such as vinblastine (anticancer) and ajmalicine (antihypertensive). Numbering: (1) is 7-deoxyloganetic acid, (2) is 7-deoxyloganic acid, (3) is loganic acid, (4) is loganin, (5) is secologanin, (6) is tryptamine, and (7) is strictosidine. (B) Formation of loganin by CraLAMT as monitored by HPLC; elution profiles of reactions with (1) no CraLAMT, (2) no SAM and (3) CraLAMT + SAM. (C) Michaelis - Menten kinetic curve of wild type CraLAMT. Results are means ± SEM of triplicate experiments. Abbreviations: GPP for geranyl pyrophosphate, LA for loganic acid, SAM for S-adenosyl-methionine, SAH for S-adenosyl-homocysteine, LGN for loganin, Ade for adenine, Pfs for the SAH nucleosidase and Glc for glucose.

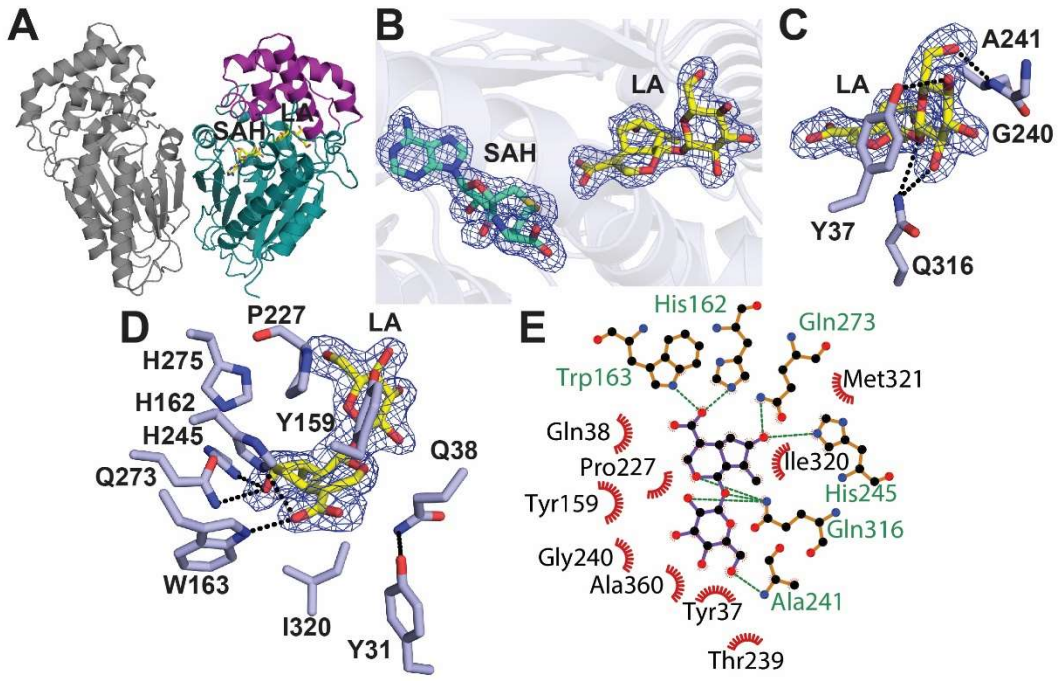

Figure 3. 2: (A) Overall structure of CraLAMT; one monomer is shown in gray. In the other monomer, the Rossmann-like domain is shown in teal and the helical domain that binds the loganic acid (LA) is shown in purple. S-adenosyl-homocysteine (SAH) and LA are shown in yellow sticks. (B) Difference Fourier map (Fo-Fc) contoured to 2.0 σ (blue) showing the bound SAH (colored in teal) and LA (colored in purple). (C, D) Stick representation showing the interactions with (C) the glucose moiety and (D) the iridoid ring of the LA. Hydrogen bonds are black dashed lines. (E) LIGPLOT [18] diagram of the LA interacting with its immediate environment. LA bonds are shown in purple, bonds within the LA environment in orange, hydrogen bonds are green dashed lines, and residues in hydrophobic contact with LA are represented by red semicircles with radiating spokes.

like domain that provides the residues for binding of the methyl donor appended to an all α-helical domain that encompasses the methyl acceptor substrate (Figure 3. 2A). The active site cavity of CraLAMT is remarkably more hydrophilic compared to the other known members of the family, and this difference reflects the differences in their substrates. Canonical SABATH methyltransferases utilize mainly bulky hydrophobic compounds, such as salicylic acid, benzoic

acid and indole-3-acetic acid while CraLAMT utilizes loganic acid, a compound with a hydrophilic glucose moiety and several polar groups.

In the CraLAMT co-crystal structure, the carboxylate group of loganic acid is engaged and oriented for methyl transfer by a pair of conserved residues, Trp163 and Gln38, and the semi-conserved His162 (Figure 3. 2D and Figure 3. 2E). Notably, the cavity of CraLAMT is different from the cavities of CbSAMT and CcXMT (Figure 3. 3), and is occupied by several polar residues that are within hydrogen bonding distance from loganic acid substrate. Additionally, a number of Tyr residues anchor the loganic acid in the correct orientation for productive methyl transfer (Figure 3. 3D). In contrast, the substrates in the structures of canonical SABATH enzymes are encircled by numerous aliphatic and aromatic residues (Figure 3. 3E and Figure 3. 3F), and are encompassed in cavities that are significantly smaller compared to the cavity of CraLAMT (Figure 3. 3A-C); the volume of solvent accessible surface for CraLAMT is ~ 2256 $Å^3$, while it is ~ 1964 $Å^3$ for CcXMT and ~ 1135 $Å^3$ for CbSAMT as calculated by CastP [39].

The CraLAMT co-crystal structure also provides insight into the selectivity towards a glycosylated and hydroxylated iridoid ring substrate (Figure 3. 2C and Figure 3. 2E). The iridoid ring is positioned by Ile320 and Tyr159 which clamp the ring on opposite sides of the ring's plane. The glucose moiety is anchored in the active site by Tyr37, which not only hydrogen bonds with the 4-hydroxyl group of glucose but also positions the glucose ring through hydrophobic interactions. In addition, Gln316 interacts with the 2-hydroxyl group and the 1-oxygen of the glucose, and the main chain amine of Ala241 is within hydrogen bonding distance from the 6-hydroxyl of glucose. The specificity of CraLAMT towards a substrate containing a hydroxylated iridoid ring [10], may be due to the positioning of residues His245 and Gln273, which are poised to interact with the hydroxyl group of the iridoid ring (Figure 3. 2D and Figure 3. 2E).

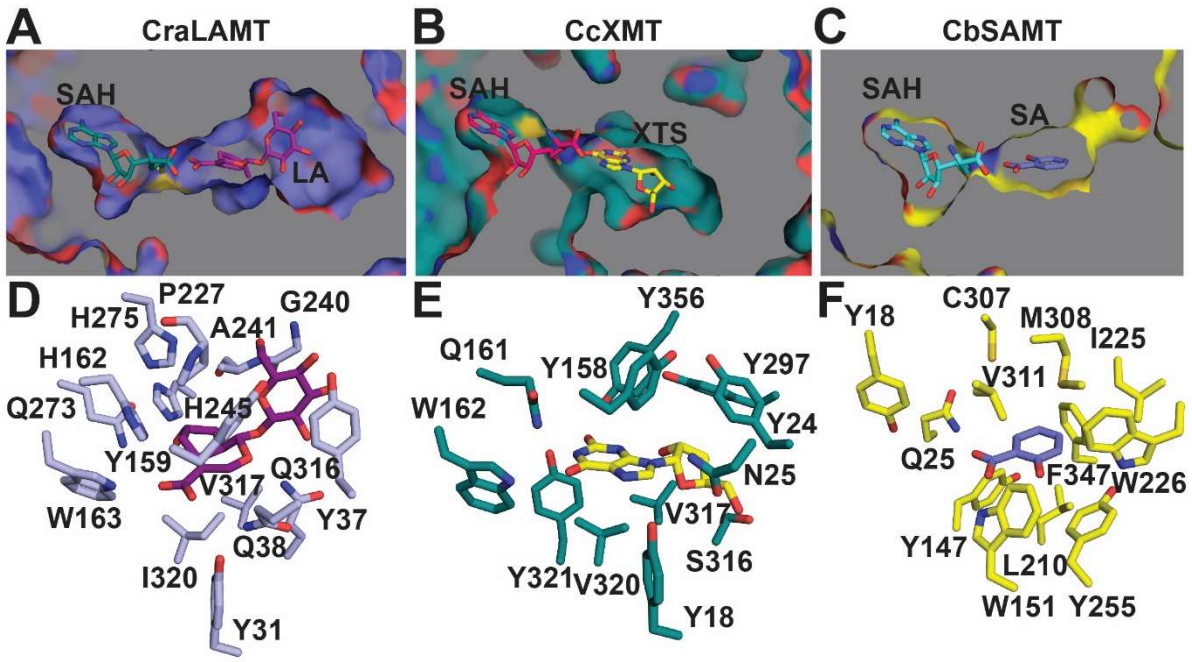

Figure 3. 3: (A, B, C) Surface cut-away diagram showing that CraLAMT (A) has a larger cavity compared to (B) the xanthosine *N*-methyltransferase from *Coffea canephora* 'robusta' (CcXMT) [38], and (C) the salicylic acid *O*-methyltransferase from *Clarkia breweri* (CbSAMT) [25]. (D, E, F) Stick representation of the residues that form the cavity where the substrates of CraLAMT (D), CcXMT (E), and CbSAMT (F) bind showing that in CraLAMT a number of hydrophilic residues form the environment of the substrate in contrast to the aliphatic and aromatic residues that line the cavities of CcXMT and CbSAMT. Abbreviations: SAH for S-adenosyl-homocysteine, LA for loganic acid, XTS for xanthosine and SA for salicylic acid.

To further investigate the significance of each of these residues on either the substrate binding or catalysis, we generated site-specific mutants and compared their activity with that of the wild type CraLAMT (Table 3. 4). The mutations Tyr159→Ala and His162→Ala abolished activity as no product formation could be detected even when 20-fold more enzyme was used. As Tyr159 clamps the iridoid ring and His162 interacts with the carboxylate group of the loganic acid, these two residues seem to be essential for proper positioning of the substrate for methylation. Mutation of any of the two residues that interact with the hydroxyl group of the iridoid ring (Gln273→Ala and His245→Ala) had an immense effect on the activity of the enzyme. The Gln273→Ala variant was only residually active (1.7%), while 20 times more enzyme had to be

used in order to detect trace product formation with the His245→Ala mutant. Finally, the mutation of Tyr37→Ala, which anchors the glucose moiety, resulted in an inactive variant. These results further demonstrate that these residues recognize and position the glycosylated and hydroxylated iridoid substrate, and are in agreement with previous studies that showed that CraLAMT had no activity towards 7-deoxy-loganic acid as well as loganetic acid [7]. Mutation of Gln316→Ala, the residue that is near the 2-hydroxyl group and the 1-oxygen of the glucose, had a significant effect on the activity (55.8%), but compared to the rest of the mutants it remained significantly active, indicating that although Gln316 plays important role in binding of the loganic acid, other interactions are more vital for recognition of the substrate.

Table 3. 4: Activity of wild type CraLAMT and mutants. Results are means ± SEM of triplicate experiments.

CraLAMT	Activity (%)	CraLAMT	Activity (%)
WT	100	**H162A**	0
Y31F	1.4 ± 0.2	**W163F**	8.1 ± 0.8
Y37A	0	**Q273A**	1.7 ± 0.5
Q38A	3.1 ± 0.5	**Q316A**	55.8 ± 9.6
Y159A	0	**H245A**	0

Finally, the mutations Trp163→Phe and Gln38→Ala had a large effect on the activity of CraLAMT (8.1% and 3.1% respectively). As Gln38 and Trp163 are near the carboxylate of the substrate, the Gln38→Ala and Trp163→Phe mutations likely compromise binding and orientation of the substrate. Although in the crystal structure of CraLAMT with SAH and loganic acid the side chain of Gln38 is not in hydrogen bonding distance from the carboxylate group, the severe effect that the Gln38→Ala mutation had on the enzyme activity suggests that Gln38 is involved in binding of the carboxylate group. This observation is in accordance with the conservation of this Gln in other carboxylic acid *O*-methyltransferases with the same fold, such as the CbSAMT (PDB

ID 1M6E) [25], the AtIAMT (PDB ID 3B5I),[14] and the fatty acid methyltransferase from *Mycobacterium marinum* (MmFAMT) (PDB ID 5F2K) [35].

Compared to the xanthosine methyltransferase (CcXMT) which also accommodates a substrate with a sugar attached to it, the active site of CraLAMT is larger and more hydrophilic (Figure 3. 2C, Figure 3. 2D and Figure 3. 3B). The cavity of XMT is lined by a number of tyrosine residues whereas the cavity of CraLAMT is comprised of a number of glutamine and histidine residues in addition to a couple of tyrosine and small aliphatic residues (Figure 3. 3D and Figure 3. 3E). Other adaptations of CraLAMT in order to incorporate a larger substrate like loganic acid include the incorporation of Pro227 at the C-terminus of beta-strand ß5 resulting in earlier termination of strands ß5, ß6 and ß7 (Figure 3. 4A and Figure 3. 4B); beta-strands ß5, ß6 and ß7 are shorter in CraLAMT compared to CcXMT and CbSAMT allowing for a bigger cavity (Figure 3. 3A-C and Figure 3. 4B).

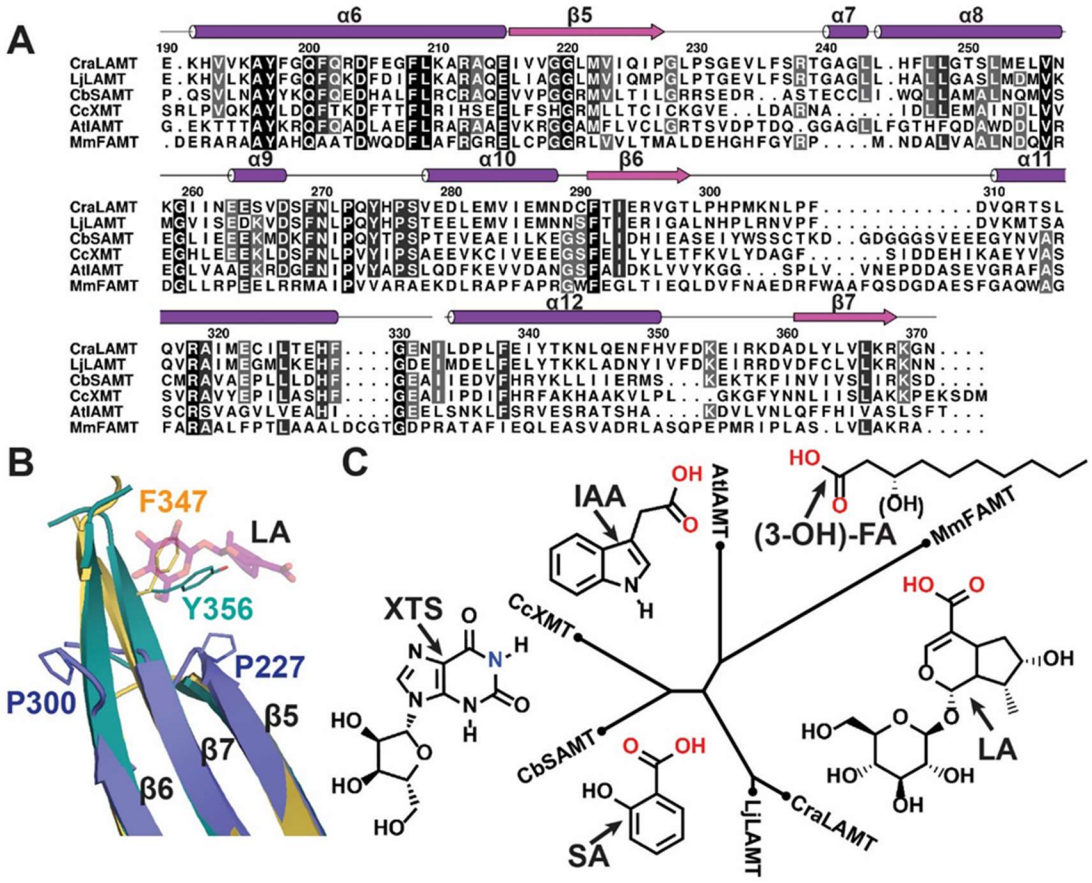

Figure 3. 4: (A) Alignment of the primary sequences of CraLAMT and five other methyltransferases that share the same fold. The secondary structure of CraLAMT is also shown. (B) Comparison of active sites of CraLAMT (in blue), CbSAMT (in yellow) [25] and CcXMT (in teal) [38]. (C) Phylogenetic tree of the methyltransferases aligned in 3A, and structures of their substrates. Abbreviations: LjLAMT for loganic acid *O*-methyltransferase from *Lonicera japonica*, CbSAMT for salicylic acid *O*-methyltransferase from *Clarkia breweri*, CcXMT for xanthosine *N*-methyltransferase from *Coffea canephora* 'robusta', AtIAMT for indole-3-acetic acid *O*-methyltransferase from *Arabidopsis thaliana*, MmFAMT for fatty acid *O*-methyltransferase from *Mycobacterium marinum* , LA for loganic acid, XTS for xanthosine, IAA for indole-3-acetic acid, (3-OH)-FA for (3-hydroxy)-fatty acid and SA for salicylic acid.

3.4 Concluding Remarks

Our biochemical and structural analysis of CraLAMT illuminates how a SABATH methyltransferase has been adapted for the methylation of a large glycosylated substrate. Although

CraLAMT shares the same fold with other members of the SABATH family, its active site cavity has evolved to accept a rather big and hydrophilic substrate. These structural data in conjunction with our mutation analysis, shed light on the molecular basis for the high specificity of CraLAMT for a glycosylated and hydroxylated iridoid ring. Finally, kinetic analysis of pure recombinant enzyme shows that pure recombinant CraLAMT is more efficient than previously reported [10], but the utilization of a solvated glycosylated substrate like loganic acid comes with a negative effect on the enzyme's K_M and, subsequently, its efficiency. This work provides a framework for further engineering studies aimed at diversification of the loganin scaffold towards the design of new classes of monoterpenoid indole alkaloids.

3.5 References

1. Moudi, M., et al., *Vinca Alkaloids.* International Journal of Preventive Medicine, 2013. **4**(11): p. 1231-1235.

2. Ziegler, J. and P.J. Facchini, *Alkaloid biosynthesis: metabolism and trafficking.* Annu Rev Plant Biol, 2008. **59**: p. 735-69.

3. Almagro, L., F. Fernandez-Perez, and M.A. Pedreno, *Indole alkaloids from Catharanthus roseus: bioproduction and their effect on human health.* Molecules, 2015. **20**(2): p. 2973-3000.

4. O'Connor, S.E. and J.J. Maresh, *Chemistry and biology of monoterpene indole alkaloid biosynthesis.* Nat Prod Rep, 2006. **23**(4): p. 532-47.

5. Yu, F. and V. De Luca, *ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in Catharanthus roseus.* Proc Natl Acad Sci U S A, 2013. **110**(39): p. 15830-5.

6. Roepke, J., et al., *Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle.* Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15287-92.

7. Miettinen, K., et al., *The seco-iridoid pathway from Catharanthus roseus.* Nat Commun, 2014. **5**: p. 3606.

8. Qu, Y., et al., *Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast.* Proc Natl Acad Sci U S A, 2015. **112**(19): p. 6224-9.

9. Brown, S., et al., *De novo production of the plant-derived alkaloid strictosidine in yeast.* Proc Natl Acad Sci U S A, 2015. **112**(11): p. 3205-10.

10. Murata, J., et al., *The leaf epidermome of Catharanthus roseus reveals its biochemical specialization.* Plant Cell, 2008. **20**(3): p. 524-42.

11. Tseng, Y.T., et al., *Loganin possesses neuroprotective properties, restores SMN protein and activates protein synthesis positive regulator Akt/mTOR in experimental models of spinal muscular atrophy.* Pharmacol Res, 2016. **111**: p. 58-75.

12. Babri, S., S. Hasani Azami, and G. Mohaddes, *Effect of Acute Administration of loganin on Spatial Memory in Diabetic Male Rats.* Adv Pharm Bull, 2013. **3**(1): p. 91-5.

13. Jiang, W.L., et al., *Effect of loganin on experimental diabetic nephropathy.* Phytomedicine, 2012. **19**(3-4): p. 217-22.

14. Li, M., et al., *The pharmacological effects of morroniside and loganin isolated from Liuweidihuang Wan, on MC3T3-E1 cells.* Molecules, 2010. **15**(10): p. 7403-14.

15. Lee, K.Y., et al., *Cognitive-enhancing activity of loganin isolated from Cornus officinalis in scopolamine-induced amnesic mice.* Arch Pharm Res, 2009. **32**(5): p. 677-83.

16. Madyastha, K.M., et al., *S-Adenosyl-L-methionine: loganic acid methyltransferase. A carboxyl-alkylating enzyme from Vinca rosea.* J Biol Chem, 1973. **248**(7): p. 2497-501.

17. Schmitt, H., et al., *Long-term prognosis of membranoproliferative glomerulonephritis type I. Significance of clinical and morphological parameters: an investigation of 220 cases.* Nephron, 1990. **55**(3): p. 242-50.

18. Qu, L., S. Li, and S. Xing, *Methylation of phytohormones by the SABATH methyltransferases.* Chinese Science Bulletin, 2010. **55**(21): p. 2211-2218.

19. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.

20. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-26.

21. Vonrhein, C., et al., *Data processing and analysis with the autoPROC toolbox.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 293-302.

22. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(Pt 2): p. 133-144.

23. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.

24. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(2): p. 213-221.

25. Zubieta, C., et al., *Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family.* Plant Cell, 2003. **15**(8): p. 1704-16.

26. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallographica Section D: Biological Crystallography, 2008. **64**(Pt 1): p. 61-69.

27. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D, Biological Crystallography, 2006. **62**(Pt 9): p. 1002-11.

28. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallographica Section D, Biological Crystallography, 2008. **64**(Pt 1): p. 83-9.

29. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.

30. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(4): p. 486-501.

31. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 355-67.

32. Lamzin, V.S. and K.S. Wilson, *Automated refinement of protein models.* Acta Crystallographica Section D, Biological Crystallography, 1993. **49**(Pt 1): p. 129-47.

33. Read, R.J., et al., *A new generation of crystallographic validation tools for the protein data bank.* Structure, 2011. **19**(10): p. 1395-412.

34. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(Pt 1): p. 12-21.

35. Petronikolou, N. and S.K. Nair, *Biochemical Studies of Mycobacterial Fatty Acid Methyltransferase: A Catalyst for the Enzymatic Production of Biodiesel.* Chem Biol, 2015. **22**(11): p. 1480-90.

36. Choi-Rhee, E. and J.E. Cronan, *A nucleosidase required for in vivo function of the S-adenosyl-L-methionine radical enzyme, biotin synthase.* Chem Biol, 2005. **12**(5): p. 589-93.

37. Zhao, N., et al., *Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family.* Plant Physiol, 2008. **146**(2): p. 455-67.

38. McCarthy, A.A. and J.G. McCarthy, *The structure of two N-methyltransferases from the caffeine biosynthetic pathway.* Plant Physiol, 2007. **144**(2): p. 879-89.

39. Dundas, J., et al., *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W116-8.

CHAPTER 4: INSIGHTS INTO THE MECHANISMS OF INHIBITION AND RESISTANCE BY THE PHOSPHONATE AMINO ACID (Z)-L-2-AMINO-5-PHOSPHONO-3-PENTENOIC ACID (APPA)[3]

4.1 Introduction

Phosphonates are natural products containing a carbon-phosphorous (C-P) bond [1]. These C-P containing molecules inhibit enzymes which utilize phosphate esters and carboxylic acids by mimicking their substrates [1]. At a time when drug-resistant pathogens pose serious threat to human health, phosphonate natural products provide a pool of potential therapeutic candidates. Two such phosphonate containing compounds are rhizocticin and plumbemycin.

Rhizocticin and plumbemycin are di- or tripeptides that contain the same phosphonic amino acid at their C-terminus, but differ at their N-terminus (Figure 4. 1A) [2, 3]. This unusual amino acid is (Z)-L-2-amino-5-phosphono-3-pentenoic acid (APPA), and is the active part of rhizocticin and plumbemycin. The N-terminal peptides facilitate active transport of these molecules into their target hosts. Upon entering the cell, host peptidases cleave the N-terminal residues to release APPA, which subsequently inhibits threonine synthase (ThrC) (Figure 4. 1B) and kills the cell [4]. Because *thrC* is an essential gene in bacteria and fungi, but is absent in humans, rhizocticin and plumbemycin are attractive as potential antimicrobials. Thus, understanding the mechanism of

[3] This project is in collaboration with Dr. Manuel Ortega and the W.W. Metcalf group (UIUC).

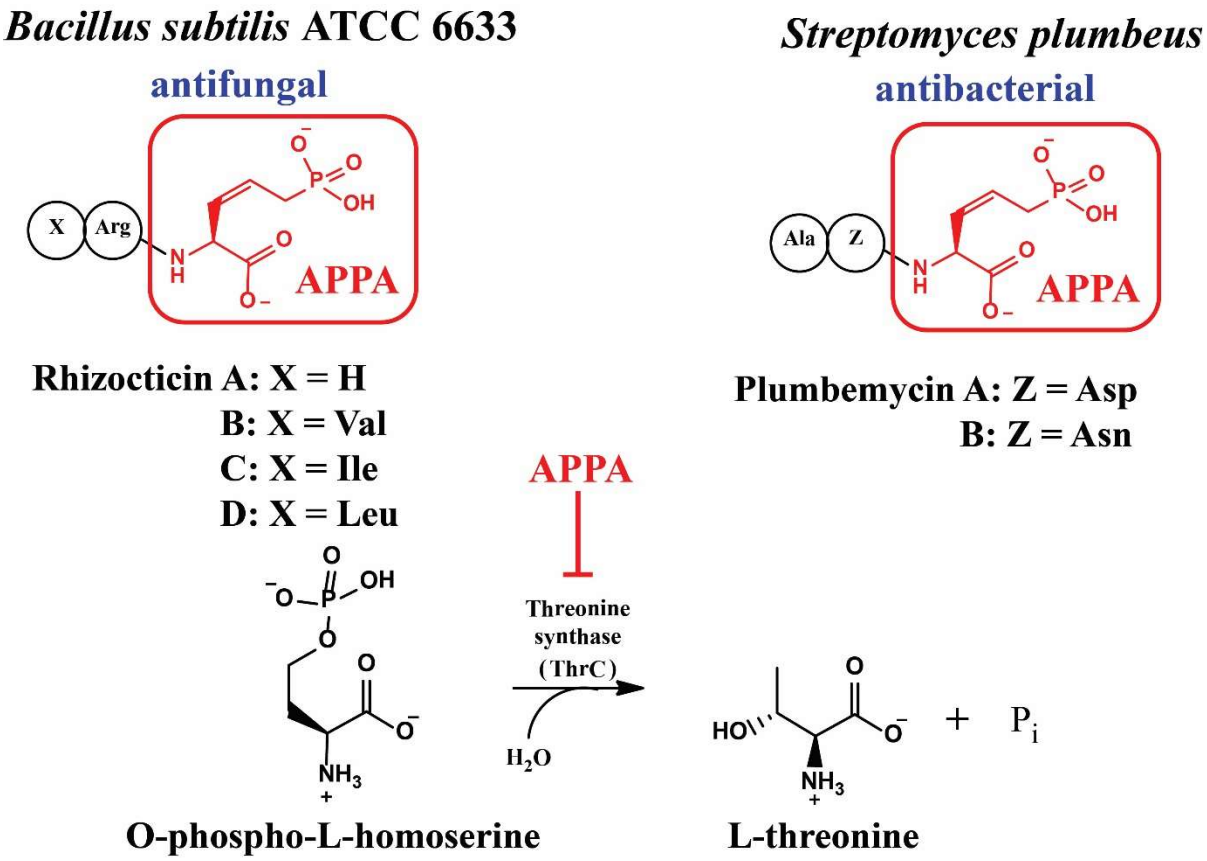

Figure 4. 1: (A) Structure of rhizocticin and plumbemycin. The (Z)-L-2-amino-5-phosphono-3-pentenoic acid (APPA) at their C-terminus is colored red. (B) Reaction catalyzed by threonine synthase. Upon release of the APPA inside the cell, APPA mimics *O*-phospho-L-homoserine and inhibits threonine synthase.

inhibition of ThrC by APPA as well as how the producing organisms overcome APPA toxicity and attain resistance is important.

There are two proposed mechanisms about the mode of inhibition of ThrC by APPA: 1) either APPA remains bound to the cofactor as a stable intermediate or 2) APPA forms a covalent adduct with the enzyme and the cofactor (Figure 4. 2) [5]. Currently, there is no direct evidence proving the one or the other mechanism. Results from previous experiments with ThrC from *Escherichia coli* were contradictive [5].

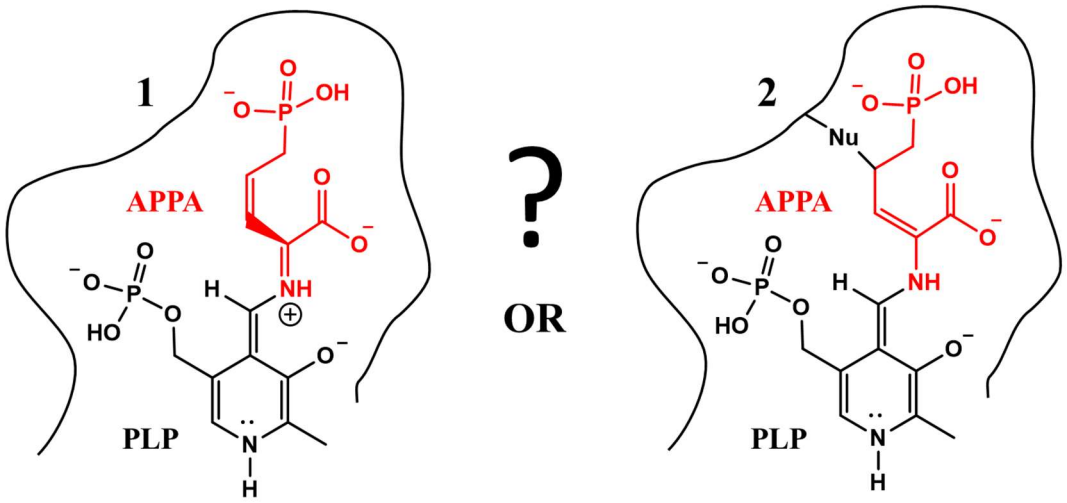

Figure 4. 2: There are two proposed mechanisms about the mode of inhibition of ThrC by APPA: 1) either APPA remains bound to the cofactor as a stable intermediate or 2) APPA forms a covalent adduct with the enzyme and the cofactor.

In addition to understanding the mechanism by which APPA inhibits threonine synthase, it is necessary to understand how the producing organisms are resistant to the inhibitor if we are interested in pursuing APPA containing compounds as potential antimicrobials. Currently, it is not clear how the native producer protects itself from APPA toxicity.

In 2010, Borisova et al. identified the biosynthetic gene cluster of rhizocticin in *Bacillus subtilis ATCC6633*, and found that it contains a gene encoding for a putative threonine synthase (rhiB). They also found that *B. subtilis ATCC6633* contains a second threonine synthase gene (*thrC*) in the threonine biosynthetic operon [4]. This led to their proposal that *B. subtilis* protects itself from APPA toxicity by having a second threonine synthase (RhiB) that is insensitive to the inhibitor. To test this hypothesis, Dr. Manuel Ortega from the W.W. Metcalf group (UIUC) expressed both *B. subtilis ATCC6633* ThrC and RhiB in *E. coli* and showed that they both have threonine synthase activity (unpublished data). He also showed that APPA binds to ThrC at 1:2 molar ratio ([enzyme]:[inhibitor]), while the same is not the case for RhiB (unpublished data).

To investigate the mechanism of inhibition of ThrC by APPA, we pursued crystallographic studies of ThrC in the presence of the inhibitor. In addition, we generated a homology model for RhiB in order to identify the differences in the active sites of RhiB and ThrC that are responsible for RhiB's "resistance" to the inhibitor. Our results provide insights into the mode of inhibition of ThrC by APPA as well as how the native producer of rhizocticin is self-resistant to APPA.

To investigate the mechanism of inhibition of ThrC by APPA, we pursued crystallographic studies of ThrC in the presence of the inhibitor. In addition, we generated a homology model for RhiB in order to identify the differences in the active sites of RhiB and ThrC that are responsible for RhiB's "resistance" to the inhibitor. Our results provide insights into the mode of inhibition of ThrC by APPA as well as how the native producer of rhizocticin is self-resistant to APPA.

4.2 Materials and Methods

Chemical and Reagents

All chemical reagents were purchased from Sigma-Aldrich unless otherwise noted. All of the material used for protein production and purification were purchased from GE Healthcare.

Cloning and Site-Specific Mutagenesis

Wild type ThrC (GenBank accession number EFG91998.1 - locus ADGS01000021) and RhiB (GenBank accession number EFG91528.1 – locus ADGS01000025) genes were cloned by Dr. Manuel Ortega (W. W. Metcalf group, UIUC) into a pET28b vector with an N-terminal His_6 tag. These plasmids were used as templates for the generation of the site-specific mutants by PCR

(primers shown in Table 4. 1). The integrity of all recombinant plasmids was confirmed by sequencing (ACGT, Inc.).

Table 4. 1: Primers used for generation of ThrC and RhiB mutants.

Primer name	Sequence (5'-3')
ThrC_PET28_NheI_F	TATGGCTAGCTGGAAAGGACTTATCCATCAATATAAAG
ThrC_PET28_XhoI_R	CATTCTCGAGTCATACACGGGCCGCTCCTTTTACATATTC
ThrC_F132Y_F	GCAATTGATGGAAACTATGACGATGCGCTTAAA
ThrC_F132Y_R	TTTAAGCGCATCGTCATAGTTTCCATCAATTGC
RhiB_PET28_NheI_F	CATGGCTAGCAAACTTGTATGTATTGACTGTGGAC
RhiB_PET28_XhoI_R	CATTCTCGAGTTAATGAATCTTTAATAAACCATTCAAATCAC
RhiB_Y188F_F	AGTATTCCGGGAACTTTTGATGACGGTATCAGG
RhiB_Y188F_R	CCTGATACCGTCATCAAAAGTTCCCGGAATACT

Protein Expression and Purification of ThrC

Expression vectors bearing wild type or mutant ThrC were transformed into *E. coli* Rosetta 2(DE3) cells for heterologous protein production. A 5 ml starter culture was inoculated in 2L Luria-Bertani growth medium supplemented with 25 µg/mL chloramphenicol and 50 µg/mL kanamycin. The culture was grown at 37 °C until the absorbance at 600 nm reached 0.6 – 0.8, at which point protein production was induced by addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was then cooled to 18 °C and grown for an additional 18 h. Cells were collected by centrifugation, resuspended in Buffer A (20 mM HEPES-NaOH, pH 8.0, 500 mM NaCl, 10% glycerol), and lysed by multiple passages through a C5 Emulsiflex (Avestin) cell homogenizer. Following centrifugation of the lysate, the supernatant was applied to a 5 ml His-Trap (GE Biosciences) column that was previously equilibrated with Buffer B (20 mM HEPES-NaOH, pH 8.0, 1 M NaCl, 30 mM imidazole, 10% glycerol). The column was extensively

washed with the same buffer and elution of specifically bound protein was carried out using a gradient of increasing imidazole concentration. Fractions containing protein of the highest purity (as determined by SDS-PAGE) were pooled and 5 mM pyridoxal 5'-phosphate (PLP) was added. After 1 h of incubation with PLP, samples were passed through a PD-10 desalting column (GE Healthcare) for buffer exchange (20 mM HEPES-NaOH, pH 8.0, 300 mM NaCl, 10% glycerol).

For crystallization of ThrC, after the Ni^{2+} affinity column, the polyhistidine affinity tag was removed by overnight incubation with thrombin at 4 °C. Samples were further purified using size exclusion chromatography (Superdex Hiload 75 16/60) in a buffer of 20 mM HEPES-NaOH, pH 7.5, 300 mM KCl. All samples were concentrated using a 10,000 Da molecular weight cutoff Amicon centrifugal filters, flash frozen in liquid nitrogen and stored at -80 °C.

Protein Expression and Purification of RhiB

Expression vectors bearing wild type or mutant RhiB were transformed into *E. coli* BL21 cells containing the pGro7 plasmid for heterologous protein production. A 5 ml starter culture was inoculated in 2L Terrific Broth growth medium (12 g tryptone, 24 g yeast extract, 4 mL glycerol and 100 mL of 0.17M KH_2PO_4 and 0.72M K_2HPO_4 per Liter) supplemented with 25 µg/mL chloramphenicol and 50 µg/mL kanamycin. The culture was grown at 37 °C until the absorbance at 600 nm reached 0.8 – 1.0, at which point protein production was induced by addition of 0.3 mM IPTG and 0.8 g/L L-arabinose. The culture was then cooled to 18 °C and grown for an additional 18 h. Cells were collected by centrifugation, resuspended in Buffer A and lysed by multiple passages through a C5 Emulsiflex (Avestin) cell homogenizer. Following centrifugation of the lysate, the supernatant was applied to a 5 ml His-Trap (GE Biosciences) column that was previously equilibrated with Buffer B. Fractions containing protein of the highest purity (as

determined by SDS-PAGE) were pooled and 5 mM pyridoxal 5'-phosphate (PLP) was added. After 1 h of incubation with PLP, samples were passed through a PD-10 desalting column (GE Healthcare) for buffer exchange (20 mM HEPES-NaOH, pH 8.0, 300 mM NaCl, 10% glycerol). All samples were concentrated using a 10,000 Da molecular weight cutoff Amicon centrifugal filters, flash frozen in liquid nitrogen and stored at -80 °C.

Kinetic Characterization of ThrC and RhiB

Michaelis-Menten kinetics were performed by Dr. Manuel Ortega (W. W. Metcalf group, UIUC) using a coupled assay for the continuous monitoring of inorganic phosphate (P_i) release by the enzyme [6]. Briefly, in the presence of inorganic phosphate, 2-amino-6-mercapto-7-methylpurine riboside (MESG) is converted enzymatically by purine nucleoside phosphorylase (PNP) to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine. This conversion of MESG results in a shift in maximum absorbance from 330 nm for the substrate (MESG) to 360 nm for the product. This shift allows for continuous monitoring of P_i production by measuring the increase in absorbance at 360 nm. The following conditions were used: 50 mM HEPES-NaOH, pH 7.5, 100 μM PLP, 200 μM MESG, 0.00375 U/μL PNP, and either 0.25 μM RhiB or 0.1 μM ThrC, and various concentrations of O-phospho-L-homoserine (PHSer).

Binding of APPA

Binding of the inhibitor to the enzyme was monitored spectrophotometrically by monitoring the formation of a chromophore at ~500 nm [5]. Absorption spectra of native and mutant enzymes were recorded every 5 min, after the addition of APPA, for total 30 min in 50 mM HEPES-NaOH (pH 7.5) and at enzyme concentrations of 2 μM. The reported ratios correspond to molar ratios of

[holoenzyme]:[APPA], and were calculated after estimation of the PLP content of each enzyme preparation. The spectra were obtained using an Epoch2 (BioTek) microplate UV-visible spectrophotometer at ambient temperature.

Estimation of PLP content – holoenzyme concentration

Protein concentrations were determined by measuring the absorbance at 280 nm using the molecular weights and extinction coefficients calculated by the ProtParam tool (ExPASY server) (Table 4. 2) [7]. In order to calculate the concentration of holo-RhiB and holo-ThrC in the protein preparations, the absorbance at 388 nm before (free PLP) and after (total PLP) the addition of 0.2 M NaOH was measured, as previously described [8-11]. The difference between the two values corresponds to the PLP concentration that is bound to the enzyme (holoenzyme concentration). Free PLP has a maximum absorbance at 388 nm with an extinction coefficient of 6600 M^{-1} cm^{-1} in 0.1 M NaOH [11].The ε_{388} of free PLP in 0.2 M NaOH was measured by using 40 μM and 80 μM PLP, and was found to be ~ 6525 M^{-1} cm^{-1}. All measurements were performed in triplicates.

Table 4. 2: Extinction coefficients and molecular weights as calculated by ProtParam (ExPASY server) and used for determination of the concentrations of wild type and mutant ThrC and RhiB.

Enzyme	**ε_{280} ($M^{-1}cm^{-1}$) x 10^3**	**MW (kDa)**
ThrC_WT	28.4	39.72
ThrC_F132Y	29.9	39.73
RhiB_WT	53.3	50.08
RhiB_Y188F	51.8	50.07

Crystallization of ThrC

Initial crystallization conditions were determined by the sparse matrix sampling method using commercial screens. Crystals of ThrC-APPA were grown using the hanging drop vapor diffusion method. Briefly, 1 µl of protein at 4 mg/ml concentration was incubated with 15 molar equivalents of APPA (1.5 mM) for 1.5 h at ambient temperature, mixed with 1 µl of precipitant solution (0.03 M $MgCl_2$ hexahydrate, 0.03 M $CaCl_2$ dihydrate, 18% v/v PEG 550 MME, 9% w/v PEG 20,000, 0.1 M MES/imidazole, pH 6.5), and equilibrated over a well containing the same precipitant solution at 9 °C. The crystals were sequentially soaked in precipitant solution supplemented with increasing concentrations of 20 and 30% ethylene glycol prior to vitrification by direct immersion in liquid nitrogen.

The ThrC-Ala/ThrC-PLP crystals were pooled from the initial screening trays. Briefly, 0.2 µl of protein at 6 mg/ml was mixed with 0.2 µl of MORPHEUS crystallization screen [12] condition H1 (0.2 M sodium L-glutamate, 0.2 M DL-alanine, 0.2 M glycine, 0.2 M DL-lysine HCl, 0.2 M DL-serine, 10% w/v PEG 20 000, 20% v/v PEG MME 550, 0.1 M MES/imidazole, pH 6.5) using the sitting drop vapor diffusion method, and equilibrated over a well containing the same precipitant solution at 9 °C. The crystals were sequentially soaked in precipitant solution supplemented with increasing concentrations of 20 and 30% ethylene glycol prior to vitrification by direct immersion in liquid nitrogen.

Data Collection, Phasing and Structure Determination

X-ray diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT), Sector 21, Argonne National Laboratory. All data were indexed, integrated and scaled using AutoProc [13]. Both structures were determined by molecular replacement as implemented in the

Phenix program suite [14]; for the ThrC-PLP/ThrC-Ala structure, the coordinates of threonine synthase from *Aquifex aeolicus* Vf5 (59% sequence identity, 77% sequence similarity, 97% cover, PDB ID: 2ZSJ) were used as search model, while for the ThrC-APPA complex the refined coordinates of ThrC-PLP/ThrC-Ala were used. In both cases, the resultant solutions were subsequently used as starting models for several rounds of automated model building using Phenix Autobuild [15] and Buccaneer [16-18], followed by rounds of manual rebuilding using Coot [19], and refinement using either Phenix Refine [20] or REFMAC5 [21]. Water molecules were added during refinement with Phenix Refine, and confirmed by manual inspection. In all cases, the quality of the in-progress model was routinely monitored using both the free R factor [22] and MolProbity [23] for quality assurance.

Homology Modeling of RhiB

The structure of the RhiB protein was modeled from its primary amino acid sequence by using the structure-based homology modeling tools Phyre2 [24] and SWISS-MODEL [25, 26]. For the Phyre2 model, the intensive mode of Phyre2 was selected and a model was generated based on several templates. For the SWISS-MODEL generated models, several different templates were used (PDB IDs: 2C2G, 3AEY, 3AEX) including the ThrC-APPA and ThrC-PLP since a conformational change that closes the active site occurs upon binding of the substrate. All models were superimposed with each other and with the ThrC-APPA structure, and were used to identify differences between the active sites of ThrC and RhiB. The active site of all models was identical. Differences were observed between the models generated by apo-templates and the models generated by templates in a closed conformation. The model generated by SWISS-MODEL using the ThrC-APPA structure as template was finally selected and is shown in the figures.

Inhibition Assays – ^{31}P NMR

Enzyme inhibition assays were carried out at 30 °C in 50 mM HEPES-NaOH (pH 7.5). In a final volume of 400 µL, 0.5 µM holo-ThrC wild type and mutant was pre-incubated with 2.5 µM APPA for 1 h at 30 °C, and then 1.6 mM PHSer (~5 x K_M) was added. For RhiB wild type and mutant, 3 µM holoenzyme was pre-incubated with 15 µM or 60 µM APPA for 1 h at 30 °C, and then 4 mM PHSer (~5 x K_M) was added. After 30 min incubation of the enzymes with PHSer, the samples were passed through 10,000 Da molecular weight cutoff Amicon centrifugal filters, and 300 µL of the reactions were removed and mixed with 300 µL D_2O. For quantification during ^{31}P-NMR, 0.5 mM of dimethylphosphinic acid was added as internal standard in all samples, and phosphate release was used for the calculation of product formed by each enzyme. The activity of the enzymes in the presence of APPA was expressed as percentage of PHSer consumed in respect to the corresponding reaction that contained no APPA (considered as 100%). Results are means ± SEM of triplicate experiments. NMR analyses were carried out in the Carl R. Woese Institute for Genomic Biology Core Facilities on a 600MHz NMR funded by NIH grant number S10-RR028833.

4.3 Results and Discussion

Enzyme Activity and Michaelis-Menten Kinetics

The threonine synthase activity of ThrC and RhiB from *B. subtillis* ATCC6633 was confirmed by Dr. Manuel Ortega (W. W. Metcalf group, UIUC) by detection of threonine utilizing ^{1}H-NMR (unpublished data). In addition, Dr. Ortega obtained the Michaelis-Menten parameters for each of these enzymes; the housekeeping threonine synthase, ThrC, was found to be ~7 times

more efficient than RhiB, the enzyme which we hypothesized to be insensitive to APPA (unpublished data).

Binding of APPA

Binding of the inhibitor to the enzyme was monitored spectrophotometrically by monitoring the formation of a chromophore at ~500 nm, as previously reported with experiments performed with the threonine synthase from *E. coli* [5]. The initial binding experiments were performed by Dr. Ortega. Briefly, ThrC and RhiB were incubated with inhibitor at 1:2 molar ratio ([enzyme]:[APPA]) for 15 min, and a shift in absorbance from the internal aldimine to ~519 nm was observed only in the case of ThrC (unpublished data). These data indicated that APPA binds to ThrC, but not to RhiB.

Structure of ThrC

The overall structure of ThrC is similar to other Fold Type II PLP-dependent enzymes (tryptophan synthase family); these enzymes are active as homodimers with two distinct active sites [27, 28]. The asymmetric unit is composed of two monomers, one in an open conformation with the PLP bound to the active site Lys59 through an aldimine linkage, and one in the closed conformation. In the closed conformation, a ligand was found to be bound to the cofactor even though no substrate was added. However, the crystallization condition contained a mixture of amino acids (0.2 M sodium L-glutamate, 0.2 M DL-alanine, 0.2 M glycine, 0.2 M DL-lysine HCl, 0.2 M DL-serine). After electron density swap experiments, the substrate bound to the PLP was identified as alanine. A phosphate group was also bound in the active site of the ThrC-Ala monomer. This phosphate group most likely stayed bound to the enzyme throughout the

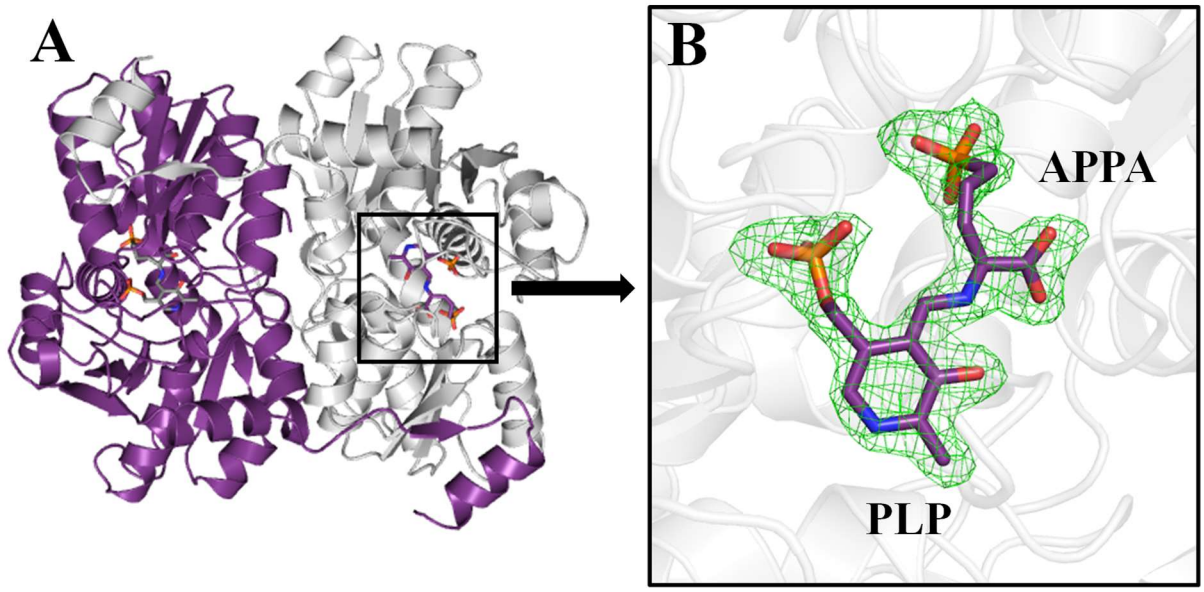

Figure 4. 3: Crystal structure of ThrC with APPA bound in the active site. (A) Ribbon representation of the ThrC-APPA homodimer. The C-terminal of each monomer crosses over to the other monomer. One monomer is colored purple and the other monomer is colored gray. (B) Difference Fourier map (Fo-Fc) contoured to 2.0 σ (green) showing the bound PLP-APPA adduct (colored in purple).

purification of ThrC. The observation of two conformations is in agreement with previous studies; upon binding of the substrate, the active site of the enzyme closes to control proton transfers and solvent accessibility during the reaction [28-30].

Mode of Inhibition

In the presence of APPA, ThrC crystallized as a dimer of dimers, and the crystals were orange, in agreement with the observed shift in the absorption spectrum of the enzyme after binding of the inhibitor [5]. The enzyme was incubated with 15 molar equivalents of APPA (1.5 mM) for 1.5 h at ambient temperature in order to make certain that the reaction with APPA had fully proceeded. The inhibitor was present in all 4 molecules in the asymmetric unit and was clearly bound to the cofactor (Figure 4. 3). As protons can be added and subtracted at different positions

during the reaction, and the covalent adduct couldn't been isolated, the exact identity of the PLP-APPA adduct in the crystal structure could not be identified. Nevertheless, it is clear from the structural data that the inhibitor remains bound to the cofactor and does not form a covalent adduct with an active site residue (mechanism 1, Figure 4. 2).

Binding of APPA to RhiB

The support of mechanism 1 (Figure 4. 2) by the crystal structure of ThrC-APPA raises the question why APPA does not bind to RhiB. Since no active site residue forms a covalent adduct with the inhibitor, then why does the inhibitor not bind to the cofactor in the active site of RhiB. To answer this question, we hypothesized that APPA can bind to RhiB, but the K_i for RhiB is

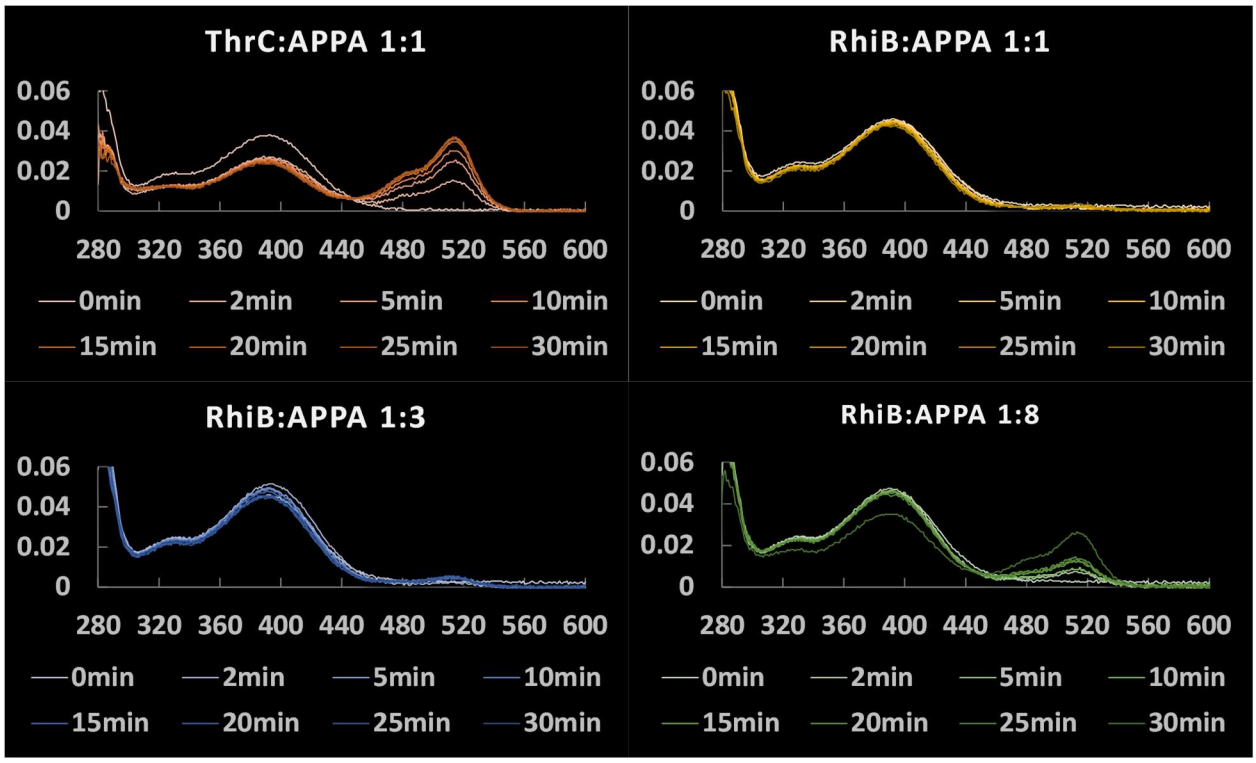

Figure 4. 4: Absorption spectra of ThrC and RhiB before and after the addition of APPA. The molar ratios indicated correspond to [holoenzyme]:[APPA]. The formation of a chromophore at ~519 nm indicates binding of the inhibitor to the cofactor. Binding of APPA to RhiB was observed at a molar ratio 1:8 and was slower compared to ThrC.

much higher than the K_i for ThrC. This lower affinity of the inhibitor towards RhiB would result in the latter being insensitive to APPA in physiological conditions (low APPA concentrations). To investigate this hypothesis, we recorded the absorption spectrum of RhiB in the presence of different concentrations of APPA, and observed binding of APPA to RhiB at a molar ratio of 1:8 ([holoenzyme]:[APPA]) (Figure 4. 4). This observation is in agreement with our hypothesis that RhiB is insensitive to APPA at physiological concentrations because the K_i for the inhibitor is much higher compared to the K_i of ThrC.

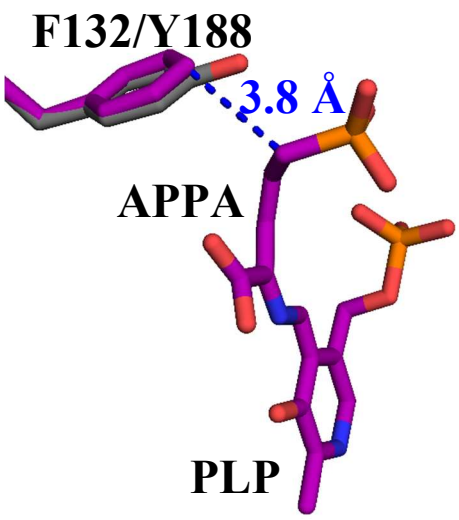

Figure 4. 5: Phe132 in ThrC (purple) is substituted by Tyr188 in RhiB (grey). Introduction of a hydroxyl group in close proximity to C5 of APPA may affect binding of the inhibitor.

In the absence of a crystal structure of RhiB, we generated a homology model for RhiB, and performed a sequence analysis of RhiB homologs in order to identify the differences in the active sites of RhiB and ThrC that are responsible for different affinities towards APPA. Comparing the two active sites, the residues that bind the cofactor were conserved. The only obvious difference was the substitution of Phe132 in ThrC by Tyr188 in RhiB (Figure 4. 5). In ThrC, Phe132 is 3.8 Å away from the C5 (carbon of the phosphonic group) of APPA, which is the position of the O5 (oxygen of the phosphate group) in PHSer. This led us to the hypothesis that introduction of a hydroxyl group in that position may affect binding of the inhibitor.

To investigate this hypothesis, we generated the ThrC_Phe132→Tyr and RhiB_Tyr188→Phe mutants and recorded the absorption spectra before and after the addition of the inhibitor (Figure 4. 6). The formation of a chromophore at ~ 514 nm indicated binding of APPA

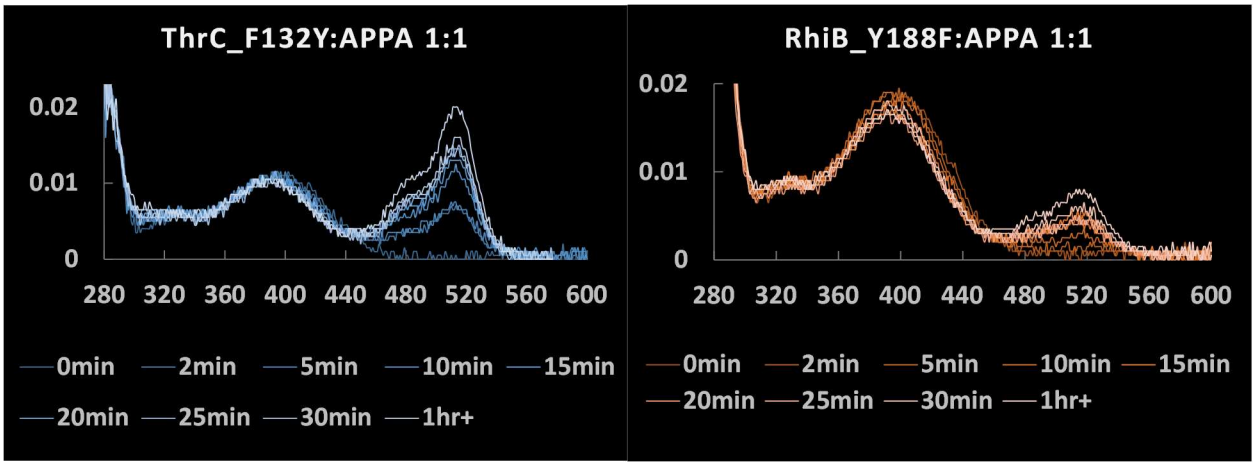

Figure 4. 6: Absorption spectra of ThrC and RhiB mutants before and after the addition of APPA. The molar ratios indicated correspond to [holoenzyme]:[APPA]. The formation of a chromophore at ~514 nm indicates binding of the inhibitor to the cofactor.

to both mutants at a molar ratio of 1:1 [holoenzyme]:[APPA]. These results indicate that Tyr188 in RhiB may play a role in the lower binding affinity for APPA, but may not be the only difference in the active site to be accounted for the different affinities.

Inhibition by APPA

In addition to the binding experiments, we performed an end-point assay and measured the activity of each enzyme variant in the absence and presence of the inhibitor (Figure 4. 7). In the case of ThrC, the activity of the wild type enzyme was about 20% in the presence of APPA while the ThrC_Phe132→Tyr variant was still 80% active. In addition, wild type RhiB was 100% active, even in the presence of 4 times more APPA compared to ThrC wild type and mutant, while the RhiB_Tyr188→Phe mutant was ~ 85% active. Substitution of Phe132 by Tyr in the APPA sensitive ThrC resulted in an enzyme that was less sensitive to the inhibitor, and substitution of Tyr188 by Phe in the APPA "insensitive" RhiB generated a variant with some sensitivity to the

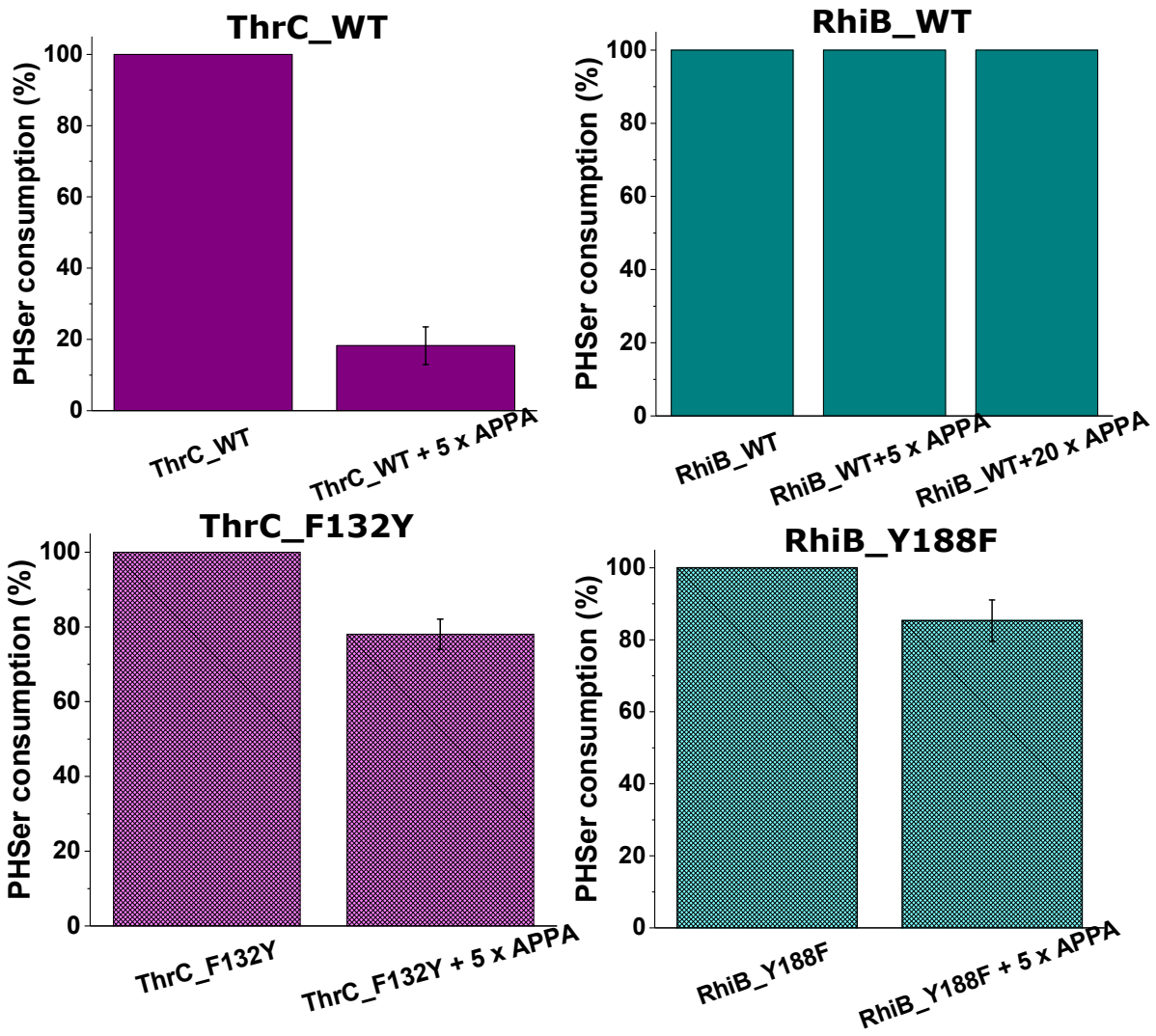

Figure 4. 7: Activity of ThrC and RhiB wild type and mutants in the absence and presence of APPA. The molar ratios indicated correspond to [holoenzyme]:[APPA].

inhibitor. These results indicate that substitution of Phe132 by Tyr in the proximity of C5 of APPA has an effect on binding of the inhibitor.

4.4 Concluding Remarks

Our structural data provide direct evidence about the mode of inhibition of threonine synthase by APPA. In conjunction with our biochemical data, they support the hypothesis that the

rhizocticin producer protects itself from APPA toxicity by having a second threonine synthase (RhiB) that is "insensitive" to the inhibitor. Our mutagenesis study suggests that the introduction of a hydroxyl group in RhiB (substitution of Phe132 by Tyr) in close proximity to the C5 of APPA has an effect on the binding of the inhibitor. However, this may not be the only difference in the active sites of ThrC and RhiB to be accounted for APPA resistance.

Further experiments that include kinetic analyses of the mutants for PHSer and rate constants of binding for APPA would further elucidate the role of Phe→Tyr substitution. Nevertheless, our data provide evidence that RhiB is a threonine synthase that is insensitive to APPA and that Tyr188 in the active site of RhiB affects the binding of the inhibitor.

4.5 References

1. Metcalf, W.W. and W.A. van der Donk, *Biosynthesis of phosphonic and phosphinic acid natural products.* Annu Rev Biochem, 2009. **78**: p. 65-94.

2. Michener, H.D. and N. Snell, *Two antifungal substances from Bacillus subtilis cultures.* Arch Biochem, 1949. **22**(2): p. 208-14.

3. Park, B.K., A. Hirota, and H. Sakai, *Studies on New Antimetabolite N-1409.* Agricultural and Biological Chemistry, 1977. **41**(1): p. 161-167.

4. Borisova, S.A., et al., *Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by Bacillus subtilis ATCC6633.* Chem Biol, 2010. **17**(1): p. 28-37.

5. Laber, B., et al., *Mechanisms of interaction of Escherichia coli threonine synthase with substrates and inhibitors.* Biochemistry, 1994. **33**(11): p. 3413-23.

6. Webb, M.R., *A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems.* Proc Natl Acad Sci U S A, 1992. **89**(11): p. 4884-7.

7. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.

8. Kulikova, V.V., et al., *Role of Arginine 226 in the Mechanism of Tryptophan Indole-Lyase from Proteus vulgaris.* Biochemistry (Moscow), 2003. **68**(11): p. 1181-1188.

9. Pennacchietti, E., et al., *Mutation of His465 alters the pH-dependent spectroscopic properties of Escherichia coli glutamate decarboxylase and broadens the range of its activity toward more alkaline pH.* J Biol Chem, 2009. **284**(46): p. 31587-96.

10. Tramonti, A., et al., *Contribution of Lys276 to the conformational flexibility of the active site of glutamate decarboxylase from Escherichia coli.* Eur J Biochem, 2002. **269**(20): p. 4913-20.

11. Peterson, E.A. and H.A. Sober, *Preparation of Crystalline Phosphorylated Derivatives of Vitamin B6.* Journal of the American Chemical Society, 1954. **76**(1): p. 169-175.

12. Gorrec, F., *The MORPHEUS protein crystallization screen.* J Appl Crystallogr, 2009. **42**(Pt 6): p. 1035-1042.

13. Vonrhein, C., et al., *Data processing and analysis with the autoPROC toolbox.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 293-302.

14. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(2): p. 213-221.

15. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.

16. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D, Biological Crystallography, 2006. **62**(Pt 9): p. 1002-11.

17. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallographica Section D, Biological Crystallography, 2008. **64**(Pt 1): p. 83-9.

18. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 235-42.

19. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(4): p. 486-501.

20. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.

21. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D, Biological Crystallography, 2011. **67**(Pt 4): p. 355-67.

22. Read, R.J., et al., *A new generation of crystallographic validation tools for the protein data bank.* Structure, 2011. **19**(10): p. 1395-412.

23. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D, Biological Crystallography, 2010. **66**(Pt 1): p. 12-21.

24. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature Protocols, 2015. **10**: p. 845.

25. Biasini, M., et al., *OpenStructure: an integrated software framework for computational structural biology.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 5): p. 701-9.

26. Bordoli, L., et al., *Protein structure homology modeling using SWISS-MODEL workspace.* Nature Protocols, 2008. **4**: p. 1.

27. Schneider, G., H. Kack, and Y. Lindqvist, *The manifold of vitamin B6 dependent enzymes.* Structure, 2000. **8**(1): p. R1-6.

28. Eliot, A.C. and J.F. Kirsch, *Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations.* Annu Rev Biochem, 2004. **73**: p. 383-415.

29. Wolfenden, R., *Enzyme catalysis: conflicting requirements of substrate access and transition state affinity.* Mol Cell Biochem, 1974. **3**(3): p. 207-11.

30. Schirch, V., et al., *The origin of reaction specificity in serine hydroxymethyltransferase.* J Biol Chem, 1991. **266**(2): p. 759-64.

APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 1

MmFAMT behaves as a homodimer in solution

To determine whether MmFAMT exists as a homodimer in solution as it does in the crystal lattice, we performed size exclusion chromatography (Superdex Hiload 75 16/60) for proteins of known molecular weights (carbonic anhydrase; ~ 29 kDa, albumin; ~ 66 kDa, and alcohol dehydrogenase; ~ 150 kDa - Sigma-Aldrich), and generated a standard curve (Figure A. 1B). The column was equilibrated in a buffer of 20 mM HEPES-KOH, pH 7.5, 300 mM KCl, and 3 mg/ml of each standard were injected and ran through the column at a flow rate of 1 mg/ml. Similarly, 3 mg/ml of MmFAMT were loaded and ran through the same column under the same conditions (buffer and flow rate). MmFAMT (~40.5 kDa) eluted at 55.82 ml (Figure A. 1A), elution volume that corresponds to a molecular weight of approximately 100 kDa. This molecular weight is in agreement with a dimer of MmFAMT with its hydration sphere.

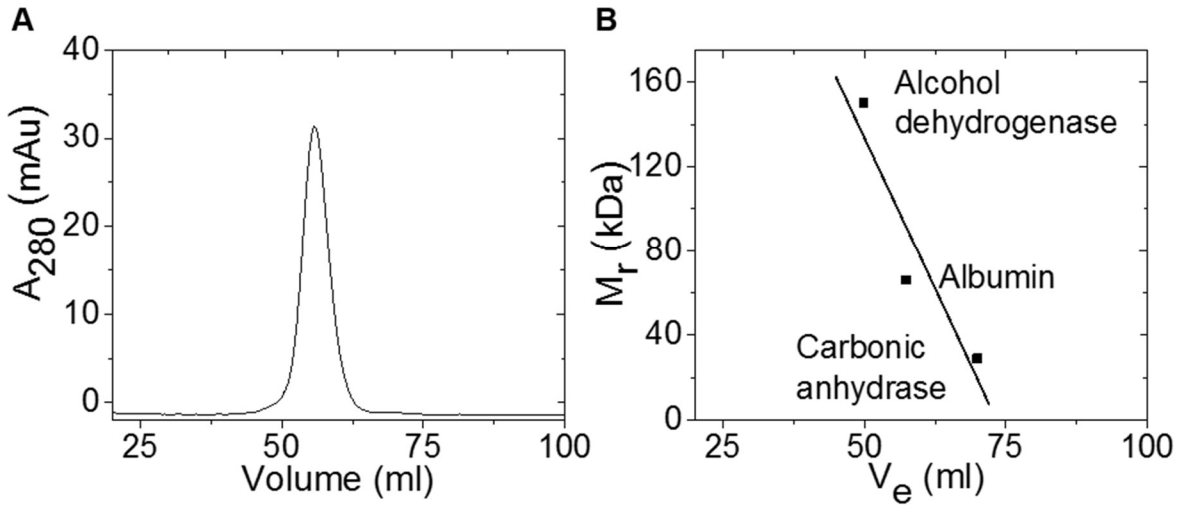

Figure A. 1: (A) Elution profile of MmFAMT using a Superdex Hiload 75 16/60 column (V_e= 55.82ml). (B) Standard curve generated by the elution volumes of protein standards; alcohol dehydrogenase is ~ 150 kDa, albumin ~ 66 kDa, and carbonic anhydrase ~ 29 kDa. MmFAMT (~40.5 kDa) eluted at 55.82 ml, elution volume that corresponds to a molecular weight of approximately 100 kDa. This molecular weight is in agreement with a dimer of MmFAMT with its hydration sphere.

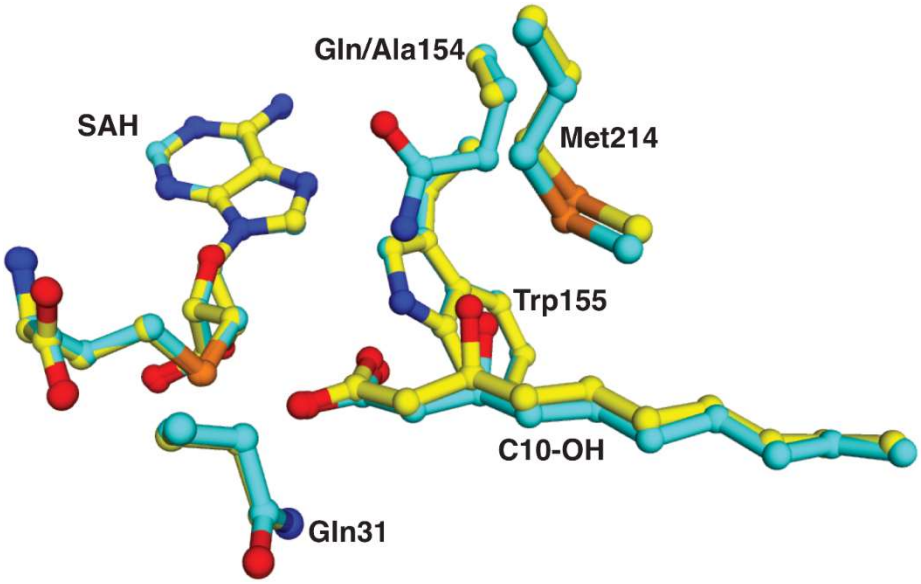

Figure A. 2: Superposition of the active sites of wild-type (cyan) and Q154A (yellow) MmFAMT showing the movement of the C10-hydroxy substrate as a consequence of the mutation.

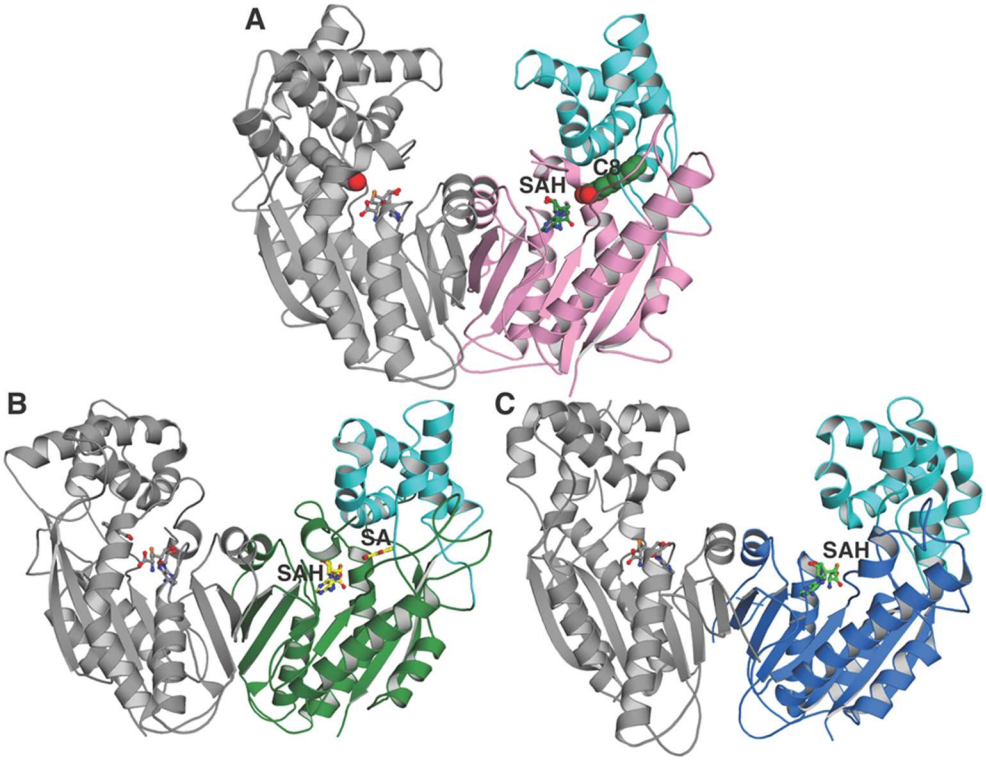

Figure A. 3: Comparison of the overall dimeric structure of (A) MmFAMT bound to octanoate (C8 shown in CPK representation) with those of (B) salicylic acid *O*-methyltransferase bound to salicylic acid (SA shown as yellow sticks) [1] and (C) indole-3-acetic acid *O*-methyltransferase [2]. One monomer is colored in gray, and the capping domain of the other monomer is colored in cyan. Bound SAH is shown as sticks.

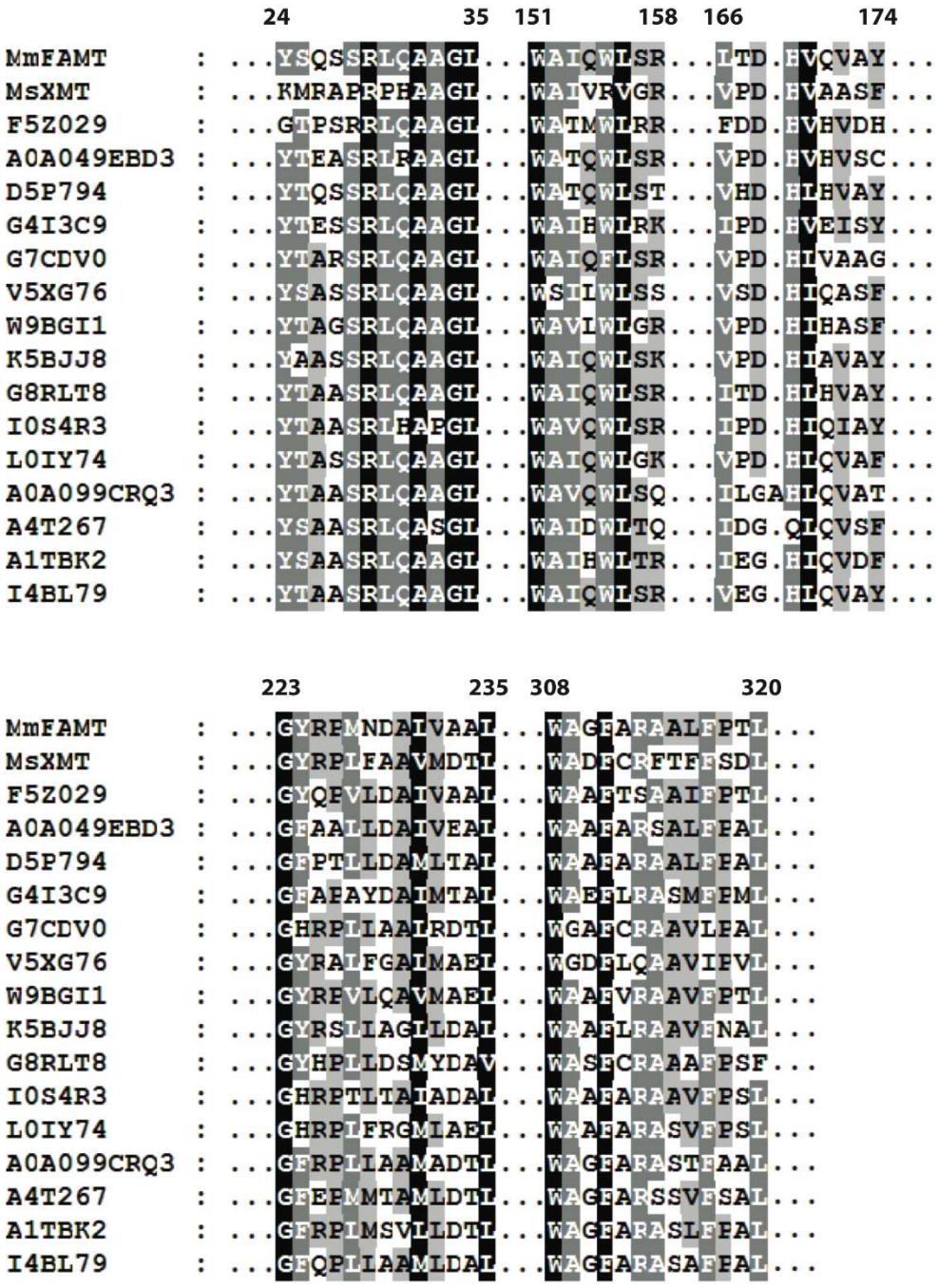

Figure A. 4: Sequence alignment of mycobacterial proteins that are 50 - 80% identical with MmFAMT and with one another. The alignment was performed by the Clustal Omega tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) using the default settings. Only parts of the sequences involved in substrate binding are shown. The UniProtKB (http://www.uniprot.org/) IDs of the sequences are listed in Table A. 3.

Table A. 1: Primers used for amplification of wild type MmFAMT and generation of mutants.

Primer name	Sequence (5'-3')
FAMT_WT_NdeI_F	GATA CATATGCCACGGGAGATCAGGCTGCC
FAMT_WT_BamHI_R	ACTAGGATCCTCAGGCGCGCTTGGCAAGCA
FAMT_Y24F_F	ATGGAAAGTGCGACCTTCAGCCAGTCGTCAAGA
FAMT_Y24F_R	TCTTGACGACTGGCTGAAGGTCGCACTTTCCAT
FAMT_Q31A_F	CAGTCGTCAAGACTAGCGGCCGCGGGCTTGTCG
FAMT_Q31A_R	CGACAAGCCCGCGGCCGCTAGTCTTGACGACTG
FAMT_Q154A_F	GGGCTGGAGTTCATGGGCGATCGCATGGCTTAGCCGAATCCCC
FAMT_Q154A_R	GGGGATTCGGCTAAGCCATGCGATCGCCCATGAACTCCAGCCC
FAMT_W155A_F	GGAGTTCATGGGCGATCCAAGCGCTTAGCCGAATCCCCGC
FAMT_W155A_R	GCGGGGATTCGGCTAAGCGCTTGGATCGCCCATGAACTCC

Table A. 2: Percentage of salicylic acid (SA) and octanoic acid (C8) converted into their *O*-methylated counterparts (Me-SA and Me-C8 respectively) by the wild type MmFAMT as determined by gas chromatography-mass spectrometry (GC/MS) in end point assays.

Reaction	%SA	%C8
1	23	17
2	18	19
3	26	17
Mean ± SED	22.3 ± 4	17.7 ± 1

Table A. 3: UniProtKB (http://www.uniprot.org/) IDs of mycobacterial sequences used for the alignment shown in Figure A. 4.

UniProt ID	Organism	Length
A0A049EBD3	*Mycobacterium avium* XTB13-223	370
A0R0D4	*Mycobacterium smegmatis* (strain ATCC 700084 / mc(2)155)	360
A0A099CRQ3	*Mycobacterium rufum*	361
A1TBK2	*Mycobacterium vanbaalenii* (strain DSM 7251 / PYR-1)	361
A4T267	*Mycobacterium flavescens* (strain ATCC 700033 / PYR-GCK)	360
B2HHT4	*Mycobacterium marinum* (strain ATCC BAA-535 / M)	368
D5P794	*Mycobacterium parascrofulaceum* ATCC BAA-614	364
F5Z029	*Mycobacterium sp.* (strain JDM601)	359
G4I3C9	*Mycobacterium rhodesiae* JS60	370
G7CDV0	*Mycobacterium thermoresistibile* ATCC 19527	364
G8RLT8	*Mycobacterium rhodesiae* (strain NBB3)	361
I0S4R3	*Mycobacterium phlei* RIVM601174	364
I4BL79	*Mycobacterium chubuense* (strain NBB4)	361
K5BJJ8	*Mycobacterium hassiacum* DSM 44199	362
L0IY74	*Mycobacterium smegmatis* JS623	350
V5XG76	*Mycobacterium neoaurum* VKM Ac-1815D	357
W9BGI1	*Mycobacterium cosmeticum*	353

References

1. Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J. P. (2003) Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *The Plant Cell* **15**, 1704-1716

2. Zhao, N., Ferrer, J. L., Ross, J., Guan, J., Yang, Y., Pichersky, E., Noel, J. P., and Chen, F. (2008) Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family. *Plant Physiology* **146**, 455-467

3. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539

4. (2015) UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204-212

APPENDIX B: SUPPORTING INFORMATION FOR CHAPTER 3

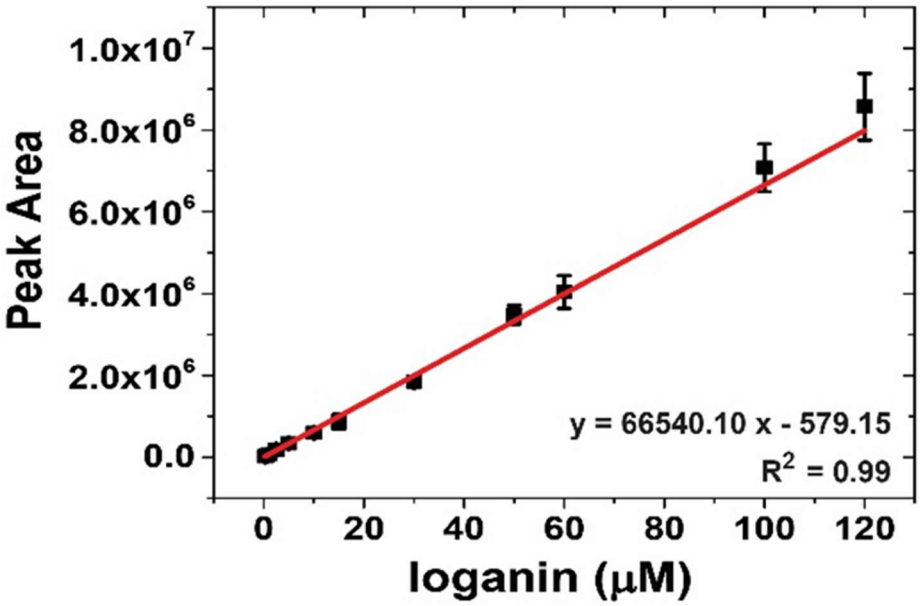

Figure B. 1: Standard curve generated by loganin standards of known concentrations, and used for the determination of the concentration of loganin produced by CraLAMT in order to perform Michaelis-Menten kinetic analysis. Results are means ± SEM of triplicate experiments.

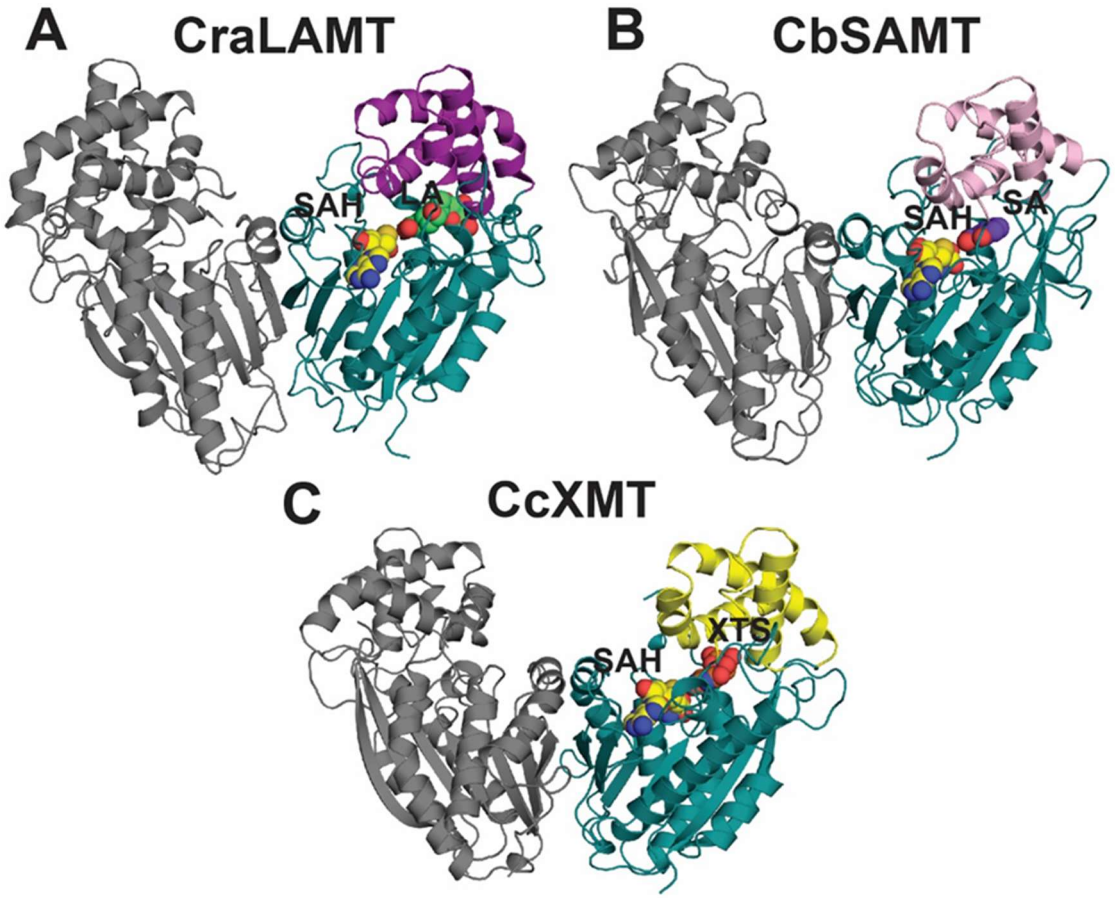

Figure B. 2: Comparison of the overall dimeric structure of (A) CraLAMT with that of (B) salicylic acid O-methyltransferase from *Clarkia breweri* (CbSAMT) [1] and (C) xanthosine N-methyltransferase from *Coffea canephora* 'robusta' (CcXMT) [2]. One monomer is colored in gray, the Rossmann-like domain is colored in teal, and the capping domain of the other monomer is shown in different colors. Bound S-adenosyl-homocysteine (SAH, shown in yellow) and substrates are shown as space-filling models. Abbreviations: LA for loganic acid, SA for salicylic acid, and XTS for xanthosine.

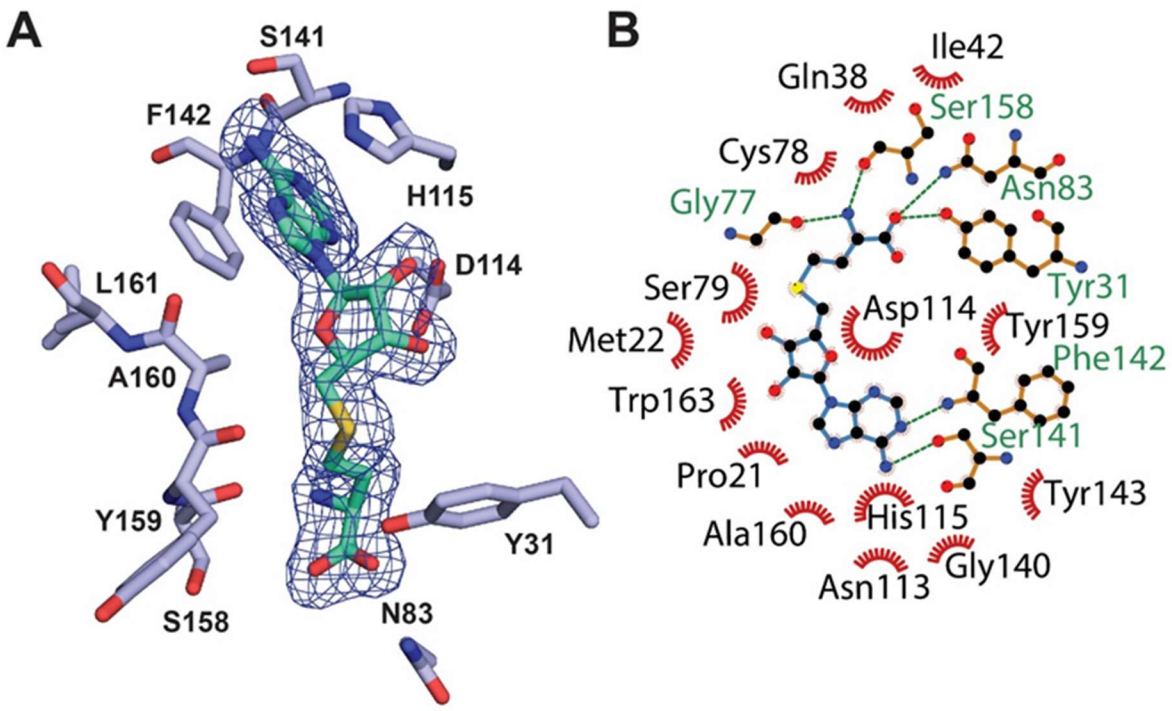

Figure B. 3: (A) Difference Fourier map (Fo-Fc) contoured to 2.0 σ (blue) showing the bound SAH (colored in teal). (B) LIGPLOT[3] diagram of the SAH interacting with its immediate environment. SAH bonds are shown in blue, bonds within the SAH environment in orange, hydrogen bonds are green dashed lines, and residues in hydrophobic contact with SAH are represented by red semicircles with radiating spokes.

Table B. 1: Primers used for amplification of wild type CraLAMT and generation of site-directed mutants.

Primer Name	Sequence (5' → 3')
CraLAMT_WT_F	TAATGGTACCATGGTTGCCACAATTGATTCCAT
CraLAMT_WT_R	GTGCTCGAGATTTCCCTTGCGTTTCAAGACAAG
CraLAMT_Y31F_F	CCACAGTTTCTCCCAAAATTCTTGCTACCAGAAAGG
CraLAMT_Y31F_R	GAATTTTGGGAGAAACTGTGGGAGTCATCACCAC
CraLAMT_Y37A_F	TCCCAAAATTCTTGCGCCCAGAAAGGAGTAATT
CraLAMT_Y37A_R	AATTACTCCTTTCTGGGCGCAAGAATTTTGGGA
CraLAMT_Q38A_F	CAAAATTCTTGCTACGCGAAAGGAGTAATTGAT
CraLAMT_Q38A_R	ATCAATTACTCCTTTCGCGTAGCAAGAATTTTG
CraLAMT_Y159A_F	TTTGCTCATTGTTCTGCTGCACTTCATTGGTTA
CraLAMT_Y159A_R	TAACCAATGAAGTGCAGCAGAACAATGAGCAAA
CraLAMT_H162A_F	TGTTCTTATGCACTTGCTTGGTTATCTAAAGTG
CraLAMT_H162A_R	CACTTTAGATAACCAAGCAAGTGCATAAGAACA
CraLAMT_W163F_F	TCTTATGCACTTCATTTCTTATCTAAAGTGCCC
CraLAMT_W163F_R	GGGCACTTTAGATAAGAAATGAAGTGCATAAGA

Table B. 2: Extinction coefficients and molecular weights as calculated by ProtParam [4] (ExPASY server) and used for determination of the concentrations of wild type CraLAMT and mutants.

CraLAMT variant	ε_{280} ($M^{-1}cm^{-1}$) x 10^3	MW (kDa)
WT	21.9	43.08
Y31F	20.4	43.07
Y37A	20.4	42.99
Q38A	21.9	43.03
Y159A	20.4	42.99
H162A	21.9	43.02
W163F	16.4	43.04
Q273A	21.9	43.03
Q316A	21.9	43.03
H245A	21.9	43.02

References

1. Zubieta, C., et al., *Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family.* Plant Cell, 2003. **15**(8): p. 1704-16.

2. McCarthy, A.A. and J.G. McCarthy, *The structure of two N-methyltransferases from the caffeine biosynthetic pathway.* Plant Physiol, 2007. **144**(2): p. 879-89.

3. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

4. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.